

General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.
- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.
- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.
- This document is paginated as submitted by the original source.
- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

"Made available under NASA sponsorship
in the interest of early and wide dis-
semination of Earth Resources Survey
Program information and without liability
for any use made thereof."

E83-10171
CR-169779

FINAL REPORT

USE OF COLLATERAL INFORMATION
TO IMPROVE LANDSAT CLASSIFICATION ACCURACIES

(E83-10171) USE OF COLLATERAL INFORMATION
TO IMPROVE LANDSAT CLASSIFICATION ACCURACIES
Final Report, 1 Apr. 1979 - 31 Mar. 1981
(California Univ.) 197 p HC AC9/MF A01

N83-17935

Unclas
CSCL 05B G3/43 00171

Alan H. Strahler,
Principal Investigator

Original photography may be purchased
from EROS Data Center
Sioux Falls, SD 57198

Geography Remote Sensing Unit
University of California
Santa Barbara, CA 93106

NASA GRANT NSG-2377

4/1/79-3/31/81

FINAL REPORT:
USE OF COLLATERAL INFORMATION
TO IMPROVE LANDSAT CLASSIFICATION ACCURACIES

By

Alan H. Strahler

This final technical report briefly summarizes the results obtained from a research grant entitled, "Use of Collateral Information to Improve Landsat Classification Accuracies," NASA Grant NSG-2377. The research supported dealt with three major issues: (1) the use of prior probabilities in maximum likelihood classification as a methodology to integrate discrete collateral data with continuously measured image density variables; (2) the use of the logit classifier as an alternative to multivariate normal classification that permits mixing both continuous and categorical variables in a single model and fits empirical distributions of observations more closely than the multivariate normal density function; and (3) the use of collateral data in a geographic information system as exercised to model a desired output information layer as a function of input layers of raster-format collateral and image database layers.

A number of existing statistical techniques can be used to merge spectral image data with collateral information, including regression, ANOVA, MANOVA, ANCOVA, MANCOVA, discriminant analysis, contingency table analysis, multivariate normal classification with or without prior probabilities, and

logit modeling. The choice of an appropriate technique depends upon the nature of the input and output variables — continuous, discrete, or categorical — and the appropriate model. The first five of these techniques assume a linear model with a fixed covariance structure for all observations. These restrictions are usually too narrow for remotely sensed data. Accordingly, we focused our attention on the two least restrictive techniques that allow the merging of continuous and categorical data: multivariate normal classification with prior probabilities and logistic classification.

In the first technique, prior information concerning the expected distribution of classes in a final classification map is incorporated into the classification through the use of prior probabilities — that is, probabilities of occurrence for classes that are based on separate, independent knowledge concerning the area to be classified. The use of prior probabilities in a classification system is sufficiently versatile to allow (1) prior weighting of output classes based on their anticipated sizes; (2) the merging of continuously varying measurements (e.g., multispectral signatures) with discrete collateral data sets (e.g., rock type, soil type); and (3) the construction of time-sequential classification systems in which an earlier classification modifies the outcome of a later one. The prior probabilities are incorporated by modifying the decision rule employed in a Bayesian-type classifier to calculate a posteriori probabilities of class membership that are based not only on the resemblance of a pixel to the class signature, but also on the weight of the class that is estimated for the final output classification. In the merging of discrete collateral information with continuous spectral values into a single classification, a set of prior probabilities (weights) is estimated for each value that the discrete collateral variable may assume (e.g., each rock type or soil type). When

multivariate normal classifications are performed, the prior probabilities appropriate to the particular pixel are used in classification. This research is summarized in the attached reprint, "The Use of Prior Probabilities in Maximum Likelihood Classification of Remotely Sensed Data," which appeared in Remote Sensing of Environment. This article is attached as Appendix A.

Logit modeling is a versatile technique that is well adapted to the classification of remotely sensed data. In this technique, which is sometimes referred to as logistic regression, a linear compound of a vector of independent variables is used to predict a logit -- the natural logarithm of ratio of the probability that a pixel belongs to a particular class to the probability that it does not. By simultaneously estimating logits for all classes for each pixel, the values of the logits can be manipulated to yield probabilities of membership for the pixel in each of the classes. These probabilities may be retained in their own right, used to select the most likely class for a pixel, or used in a system to estimate proportions.

The logit model has two distinct advantages over conventional Gaussian classification. The first of these is that the logistic model tends to fit asymmetrical probability distributions better than does the multivariate normal approximation. Second, the logit classifier allows the modeling of a probability density as a function of any type of variable in a linear or curvilinear model. Thus, it is considerably less restrictive than multivariate normal classification. The logit modeling work is presented in three papers forming Appendices B, C and D to this report. These are, "Incorporating Collateral Data in Landsat Classification and Modeling Procedures" (from the Proceedings of the Fourteenth International Symposium on Remote Sensing of Environment), "A Logit Classifier for Multi-Image Data" (from the Proceedings of the IEEE Workshop on Picture Data Description and

Management), and "The Logit Classifier -- A General Maximum Likelihood Discriminant for Remote Sensing Applications" (from the Fifteenth International Symposium on Remote Sensing of Environment).

To implement the logit classifier, two VICAR programs were coded. The first of these, LOGPR2, accepts training sets from an input image and prints maximum likelihood values of the calibrating parameters for the training data. These parameters are then input to a second program, PRCLSQ, which calculates the probability of class membership for each pixel in an input image and produces a classification map as an output image. These programs, coded in FORTRAN IV, appear in Appendices E and F.

The last primary task supported by this grant was the development of a geographic information system using both collateral and image data in a nontrivial application in order to explore the operational difficulties of merging image and collateral data and using both to produce information as a desired output. This task is described in the M.A. thesis of M.A. Spanner, entitled "Soil Loss Prediction in a Geographic Information System Format," which is attached as Appendix G. In this work, soil loss due to erosion from rainfall was successfully predicted for the Santa Paula 7.5 minute quadrangle, Ventura County, California, utilizing the VICAR/IBIS image processing and geographic information system to simulate the Universal Soil Loss Equation (USLE).

The USLE requires as input rainfall, soil erodability, length of slope, slope gradient, and crop management and soil loss tolerance coefficients in order to predict annual soil loss in tons per acre per year. To provide these data, digital Landsat MSS data, USGS Digital Elevation Model topographic data, a digitized NOAA isopluvial map and digitized USDA Soil Conservation Service soil maps were used. Estimates of accuracy for the intermediate data planes

representing rainfall, soil erodability, length of slope, slope gradient, crop management and soil loss tolerance ranged from 81 to 100 percent. Incorporating slope and elevation information in the land use/land cover classification that was used to identify crops and thus crop management coefficients was noted to improve classification accuracies considerably. Watershed characteristics including slope and length of slope were effectively depicted using the DEM topographic data. Digitized soil maps were a powerful information source, allowing soil erodability and soil loss tolerance to be represented in an image format.

An accuracy analysis revealed a correlation coefficient of 0.91 between soil loss values obtained from the developed geobased model and a sample of manually derived soil loss values. In addition, the system accurately targeted soil loss problem areas for subsequent analysis by Soil Conservation Service personnel. Useful statistics from the soil maps, topographic data and land cover were also obtained. In summary, the maps and statistics produced in this application of geographic information systems technology have a high potential to provide useful information to resource managers dealing with the problem of soil loss in agricultural areas.

In addition to the articles and theses identified in the foregoing narrative, NASA Grant NSG-2377 supported small portions of other theses and publications. The list of publications below includes these materials as well.

List of Publications

Maynard, Paul F., 1981, The Logit Classifier: A General Maximum Likelihood Discriminant for Remote Sensing Applications. M.A. Thesis, University of California, Santa Barbara.

Maynard, P.F., and A.H. Strahler, 1981, The Logit Classifier: A General Maximum Likelihood Discriminant for Remote Sensing Applications: Proc. Fifteenth Int. Symp. on Remote Sensing of Environment, pp. 213-222.

Spanner, Michael A., 1982, Soil Loss Prediction in a Geographic Information System Format. M.A. Thesis, University of California Santa Barbara.

Stow, Douglas A., 1978, Analysis of Landsat and Digital Terrain Tape Data in a Geobase Information Systems Context: Ventura County Study Area. M.A. Thesis, University of California, Santa Barbara.

Strahler, A.H., 1981, The Use of Prior Probabilities in Maximum Likelihood Classification of Remotely Sensed Data: Remote Sensing of Environment, vol. 10, pp. 135-163.

Strahler, A.H., J.E. Estes, P.F. Maynard, F.C. Mertz, and D.A. Stow, 1980, Incorporating Collateral Data in Landsat Classification and Modeling Procedures: Proc. Fourteenth Int. Symp. on Remote Sensing of the Environment, pp. 1009-1026.

Strahler, A.H. and P.F. Maynard, 1980, A Logit Classifier for Multi-Image Data: Proc. of the Workshop on Picture Data Description and Management, August 17-28, 1980, Asilomar, California, IEEE Computer Society 80CH1530-5, pp. 18-26.

Strahler, A.H. and P.F. Maynard, 1981, Combining Landsat and Collateral Data in Classification and Modeling Procedures: Landsat 81 -- Proc. Second Australasian Remote Sensing Conference, Canberra, Australis, pp. 6.11.1.-6.11.6.

A P P E N D I X A

The Use of Prior Probabilities in Maximum Likelihood Classification of Remotely Sensed Data

ALAN H. STRAHLER

University of California, Santa Barbara, California

The expected distribution of classes in a final classification map can be used to improve classification accuracies. Prior information is incorporated through the use of prior probabilities—that is, probabilities of occurrence of classes which are based on separate, independent knowledge concerning the area to be classified. The use of prior probabilities in a classification system is sufficiently versatile to allow (1) prior weighting of output classes based on their anticipated sizes; (2) the merging of continuously varying measurements (multispectral signatures) with discrete collateral information datasets (e.g., rock type, soil type); and (3) the construction of time-sequential classification systems in which an earlier classification modifies the outcome of a later one. The prior probabilities are incorporated by modifying the maximum likelihood decision rule employed in a Bayesian-type classifier to calculate *a posteriori* probabilities of class membership which are based not only on the resemblance of a pixel to the class signature, but also on the weight of the class which is estimated for the final output classification. In the merging of discrete collateral information with continuous spectral values into a single classification, a set of prior probabilities (weights) is estimated for each value which the discrete collateral variable may assume (e.g., each rock type or soil type). When maximum likelihood calculations are performed, the prior probabilities appropriate to the particular pixel are used in classification. For time-sequential classification, the prior classification of a pixel indexes a set of appropriate conditional probabilities reflecting either the confidence of the investigator in the prior classification or the extent to which the prior class identified is likely to change during the time period of interest.

Introduction

In the past ten years, maximum likelihood classification has found wide application in the field of remote sensing. Based on multivariate normal distribution theory, the maximum likelihood classification algorithm has been in use for applications in the social sciences since the late 1940s. Providing a probabilistic method for recognizing similarities between individual measurements and pre-defined standards, the algorithm found increasing use in the field of pattern recognition in the following decades (Chow, 1957; Sebestyen, 1962; Nilsson, 1965). In remote sensing, the development of multispectral scanning technology to produce layered multispectral digital images of land areas from aircraft or

spacecraft provided the opportunity to use the maximum likelihood criterion in producing thematic classification maps of large areas for such purposes as land-use/land-cover determination and natural cultivated land inventory (Schell, 1972; Reeves et al., 1975).

In the last decade, research on the general use of classification algorithms in remote sensing has centered in two areas: (1) computational improvements in evaluating maximum likelihood and discriminant function decision rules; and (2) the use of various unsupervised clustering algorithms to extract repeated or commonly occurring measurement vectors which are characteristic of a particular multispectral scene. Computational improvements have included such developments as look-up table schemes (Schlien

and Smith, 1975) to reduce repeated calculation, and hybrid classifiers (Addington, 1975) which use parallelepiped algorithms (Goodenough and Schlien, 1974) first, then turn to maximum likelihood computation to resolve ambiguities. Although important for small image processing systems, further computational improvements will become less and less cost effective as real-time computational costs continue to fall due to the development of fourth- and fifth-generation computer hardware systems.

Unsupervised methods rely on clustering measurement vectors according to some set of distance, similarity, or dispersion criteria. Many clustering heuristics have been devised and applied in image processing. Dubes and Jain (1976) provide a review and comparative analysis of a number of techniques which are commonly applied in pattern recognition. However, as Kendall (1972, p. 291) points out, clustering is a subjective matter to which little probabilistic theory is applicable. No clustering algorithms have as yet come to the fore which can incorporate prior knowledge in a formal fashion (except for the use of *a priori* starting vectors in interactive clustering) with an expected increment in class identification accuracy produced by the use of this additional information. However, recent developments involving guided clustering and automated labeling of unsupervised clusters blur the distinction between supervised and unsupervised techniques. Future work may well produce a continuum of intergrading methods from which a user can select a mix appropriate to the spatial, spectral, and temporal resolution of the data in hand and information output desired.

The purpose of this paper is to show how the use of prior information about the expected distribution of classes in a final classification map can be used in several different models to improve classification accuracies. Prior information is incorporated through the use of prior probabilities—that is, probabilities of occurrence of classes which are based on separate, independent knowledge concerning the area to be classified. Used in their simplest form, the probabilities weight the classes according to their expected distribution in the output dataset by shifting decision space boundaries to produce larger volumes in measurement space for classes that are expected to be large and smaller volumes for classes expected to be small.

The incorporation of prior probabilities into the maximum likelihood decision rule can also provide a mechanism for merging continuously measured observations (multispectral signatures) with discretely measured collateral variables such as rock type or soil type. As an example, consider an area of natural vegetation underlain by two distinctive rock types, each of which exhibits a unique mix of vegetation classes. Two sets of prior probabilities can be devised, one for each rock type, and the classifier can be modified to use the appropriate set of prior probabilities contingent on the underlying rock type. In this way, the classification process can incorporate discrete collateral information into the decision rule through a model contingent on an external conditioning variable. The method can also be extended to include two or more such discrete collateral datasets; the number is limited only by the ability to estimate the required sets of prior proba-

bilities. Thus, prior probabilities provide a powerful mechanism for merging collateral datasets with multispectral images for classification purposes.

Another application of prior probabilities contingent upon a collateral dataset allows temporal weighting in a time-sequential classification system. As an example, consider distinguishing between two crop types which, through differing phenologies, can be easily separated early in the season but are confused later on in the growing period. Through the use of prior probabilities, a midsummer classification can "look backward" to a spring classification to resolve ambiguity. Thus, winter wheat could be separated from spring wheat at midseason by its distinctive early spring signature. This use of temporal information provides an alternative to the calculation of transformed vegetation indexes (TVIs) and comparable procedures (Richardson and Wiegand, 1977) in the identification of crops with multitemporal images (Rouse et al., 1973). Such a time-sequential classification system could also be used to monitor land use change. In this case, a Markov-type predictive model is used directly to set prior probabilities based on patterns of change shown in an area.

Review of Maximum Likelihood Classification

To understand the application of prior probabilities to a classification problem, we must first briefly review the mathematics of the maximum likelihood decision rule. For the multivariate case, we assume each observation \mathbf{X} (pixel) consists of a set of measurements on p variables (channels). Through some external

procedure, we identify a set of observations which correspond to a class—that is, a set of similar objects characterized by a vector of means on measurement variables and a variance-covariance matrix describing the interrelationships among the measurement variables which are characteristic of the class. Although the parametric mean vector and dispersion matrix for the class remain unknown, they are estimated by the sample means and dispersion matrix associated with the object sample.

Multivariate normal statistical theory describes the probability that an observation \mathbf{X} will occur, given that it belongs to a class k , as the following function:

$$\Phi_k(\mathbf{X}_i) = (2\pi)^{-1/2p} |\Sigma_k|^{-1/2} \times e^{-1/2(\mathbf{X} - \mu_k)' \Sigma_k^{-1} (\mathbf{X} - \mu_k)}. \quad (1)$$

(Table 1 presents an explanation of the symbols used in this and other expressions.) The quadratic product

$$\chi^2 = (\mathbf{X} - \mu_k)' \Sigma_k^{-1} (\mathbf{X} - \mu_k) \quad (2)$$

can be thought of as a squared distance function which measures the distance between the observation and the class mean as scaled and corrected for variance and covariance of the class. It can be shown that this expression is a χ^2 variate with p degrees of freedom (Tatsuoka, 1971).

As applied in a maximum likelihood decision rule, expression (1) allows the calculation of the probability that an observation is a member of each of k classes. The individual is then assigned to the class for which the probability value is greatest. In an operational context, we substitute observed means, variances, and

TABLE I Notation

TERM	DEFINITION
p	Number of measurement variables used to characterize each object or observation.
\mathbf{X}	A p -dimensional random vector.
\mathbf{X}_i	Vector of measurements on p variables associated with the i th object or observation; $i = 1, 2, \dots, N$.
$P(\mathbf{X}_i)$	Probability that a p -dimensional random vector \mathbf{X} will take on observed values \mathbf{X}_i .
ω_k	Member of the k th set of classes ω ; $k = 1, 2, \dots, K$.
v_j	Member of the j th set of states for a conditioning variable v ; $j = 1, 2, \dots, J$.
$P(\omega_k)$	Probability that an observation will be a member of class ω_k ; prior probability for class ω_k .
$P(v_j)$	Probability that an observation will be associated with state j of conditioning variable v ; prior probability for state v_j .
$P(\omega_k \mathbf{X}_i)$	Probability that an observation is a member of class ω_k given that measurement vector \mathbf{X}_i is observed.
$\Phi_k(\mathbf{X}_i)$	Probability density value associated with observation vector \mathbf{X}_i as evaluated for class k .
μ_k	Parametric mean vector associated with the k th class.
\mathbf{m}_k	Mean vector associated with a sample of observations belonging to the k th class; taken as an estimator of μ_k .
Σ_k	Parametric p by p dispersion (variance-covariance) matrix associated with the k th class.
\mathbf{D}_k	p by p dispersion matrix associated with a sample of observations belonging to the k th class; taken as an estimator of Σ_k .

covariances and use the log form of expression (1)

$$\begin{aligned} \ln[\Phi_k(\mathbf{X}_i)] &= -\frac{1}{2}p\ln(2\pi) - \frac{1}{2}\ln|\Sigma_k| \\ &\quad - \frac{1}{2}(\mathbf{X}_i - \mathbf{m}_k)' \mathbf{D}_k^{-1} (\mathbf{X}_i - \mathbf{m}_k). \end{aligned} \quad (3)$$

Since the log of the probability is a monotonic increasing function of the probability, the decision can be made by comparing values for each class as calculated from the right hand side of this equation. This is the decision rule that is used in the currently distributed versions of LARSYS and VICAR, two image processing program systems authored respectively by the Laboratory for Applications of Remote Sensing at Purdue Uni-

versity and the Jet Propulsion Laboratory of California Institute of Technology at Pasadena. A simpler decision rule, R_1 , can be derived from expression (3) by eliminating the constants (Tatsuoka, 1971):

R_1 : Choose k which minimizes

$$F_{1,k}(\mathbf{X}_i) = \ln|\mathbf{D}_k| + (\mathbf{X}_i - \mathbf{m}_k)' \mathbf{D}_k^{-1} (\mathbf{X}_i - \mathbf{m}_k). \quad (4)$$

The Use of Prior Probabilities in the Decision Rule

The maximum likelihood decision rule can be modified easily to take into account prior probabilities which describe how likely a class is to occur in the

population of observations as a whole. The prior probability itself is simply an estimate of the proportion of the objects which will fall into a particular class. These prior probabilities are sometimes termed "weights," since the modified classification rule will tend to weigh more heavily those classes with higher prior probabilities.

Prior probabilities are incorporated into the classification through manipulation of the Law of Conditional Probability. To begin, we define two probabilities: $P(\omega_k)$, the probability that an observation will be drawn from class ω_k ; and $P(X_i)$, the probability of occurrence of the measurement vector X_i . The Law of Conditional Probability states that

$$P(\omega_k|X_i) = \frac{P(\omega_k, X_i)}{P(X_i)}. \quad (5)$$

The probability on the left-hand side of this expression will form the basis of a modified decision rule, since we wish to assign the i th observation to that class ω_k which has the highest probability of occurrence given the p -dimensional vector X_i which has been observed.

Again using the Law of Conditional Probability, we find that

$$P(X_i|\omega_k) = \frac{P(\omega_k, X_i)}{P(\omega_k)}. \quad (6)$$

In this expression, the left-hand term describes the probability that the measurement vector will take on the values X_i given that the object measured is a member of class ω_k . This probability could be determined by sampling a population of measurement vectors for observations known to be from class ω_k ; however, the distribution of such vectors

is usually assumed to be Gaussian. Note that in some cases this assumption may not hold; as an example, Brooner et al. (1971) showed significantly higher classification accuracies for crops using simulated multispectral imagery with direct estimates of these conditional probabilities than with probabilities calculated according to Gaussian assumptions. However, the use of the multivariate normal approximation is widely accepted, and, in any case, it is only under rare circumstances that sufficient data are obtained to estimate the conditional probabilities directly.

Thus, we can assume that $P(X_i|\omega_k)$ is acceptably estimated by $\Phi_k(X_i)$ and rewrite expression (6) as

$$\Phi_k(X_i) = \frac{P(\omega_k, X_i)}{P(\omega_k)}. \quad (7)$$

Rearranging, we have

$$P(\omega_k, X_i) = \Phi_k(X_i)P(\omega_k) = \Phi_k^*(X_i). \quad (8)$$

Thus, we see that the numerator of expression (5) can be evaluated as the product of the multivariate density function $\Phi_k(X_i)$ and the prior probability of occurrence of class ω_k .

To evaluate the denominator of expression (5), we note that for all k classes the conditional probabilities must sum to 1:

$$\begin{aligned} \sum_{k=1}^K P(\omega_k|X_i) &= 1 \\ &= \sum_{k=1}^K \left(\frac{\Phi_k(X_i)P(\omega_k)}{P(X_i)} \right). \end{aligned} \quad (9)$$

Therefore,

$$P\{\mathbf{X}_i\} = \sum_{k=1}^K \Phi_k(\mathbf{X}_i) P\{\omega_k\}. \quad (10)$$

Substituting Eqs. (8) and (10) into (5),

$$\begin{aligned} P\{\omega_k | \mathbf{X}_i\} &= \frac{\Phi_k(\mathbf{X}_i) P\{\omega_k\}}{\sum_{k=1}^K \Phi_k(\mathbf{X}_i) P\{\omega_k\}} \\ &= \frac{\Phi_k^*(\mathbf{X}_i)}{\sum_{k=1}^K \Phi_k^*(\mathbf{X}_i)}. \end{aligned} \quad (11)$$

The last expression, then, provides the basis for the decision rule which includes prior probabilities. Since the denominator remains constant for all classes, the observation is simply assigned to the class for which $\Phi_k^*(\mathbf{X}_i)$, the product of $\Phi_k(\mathbf{X}_i)$ and $P\{\omega_k\}$, is a maximum. In its simplest form, this decision rule can be stated as:

R_2 : Choose k which minimizes

$$\begin{aligned} F_{2,k}(\mathbf{X}_i) &= \ln|\mathbf{D}_k| \\ &\quad + (\mathbf{X}_i - \mathbf{m}_k)' \mathbf{D}_k^{-1} (\mathbf{X}_i - \mathbf{m}_k) \\ &\quad - 2 \ln P\{\omega_k\}. \end{aligned} \quad (12)$$

(Tatsuoka and Tiedeman, 1954).

It is important to understand how this decision rule behaves with different prior probabilities. If the prior probability $P\{\omega_k\}$ is very small, then its natural logarithm will be a large negative number; when multiplied by -2 , it will become a large positive number and thus $F_{2,k}$ for such a class will never be minimal. Therefore, setting a very small prior probability will effectively remove a class from the output classification. Note that this effect will occur even if the observa-

tion vector \mathbf{X}_i is coincident with class mean vector \mathbf{m}_k . In such a case, the quadratic product distance function $(\mathbf{X}_i - \mathbf{m}_k)' \mathbf{D}_k^{-1} (\mathbf{X}_i - \mathbf{m}_k)$ goes to zero, but the prior probability term $-2 \ln P\{\omega_k\}$ can still be large. Thus, it is entirely possible that the observation will be classified into a different class, one for which the distance function is quite large.

As the prior probability $P\{\omega_k\}$ becomes large and approaches 1, its logarithm will go to zero and $F_{2,k}$ will approach $F_{1,k}$ for that class. Since this probability and all others must sum to one, however, the prior probabilities of the remaining classes will be small numbers and their values of $F_{2,k}$ will be greatly augmented. The effect will be to force classification into the class with high probability. Therefore, the more extreme are the values of the prior probabilities, the less important are the actual observation values \mathbf{X}_i . This point is discussed in more detail in a following section.

Numerical example

A simple numerical example may clarify this modification of the maximum likelihood decision rule. For this example, we assume two classes ω_1 and ω_2 in a two-dimensional measurement space. Their means and dispersion matrices are shown below.

$$\begin{aligned} \mathbf{m}_1 &= [4 \quad 2], \quad \mathbf{m}_2 = [3 \quad 3], \\ \mathbf{D}_1 &= \begin{bmatrix} 3 & 4 \\ 4 & 6 \end{bmatrix}, \\ \mathbf{D}_2 &= \begin{bmatrix} 4 & 5 \\ 5 & 7 \end{bmatrix}. \end{aligned} \quad (13)$$

The determinants and inverses of these

matrices are

$$\begin{aligned} |\mathbf{D}_1| &= 2, \quad |\mathbf{D}_2| = 3, \\ \mathbf{D}_1^{-1} &= \begin{bmatrix} 3 & -2 \\ -2 & \frac{3}{2} \end{bmatrix}, \quad (14) \\ \mathbf{D}_2^{-1} &= \begin{bmatrix} \frac{7}{3} & -\frac{5}{3} \\ -\frac{5}{3} & \frac{4}{3} \end{bmatrix}. \end{aligned}$$

For this example, we wish to decide to which class the measurement vector $(4, 3)$ belongs.

To evaluate the probability associated with ω_1 , we first evaluate the quadratic product

$$\chi^2 = (\mathbf{X} - \mathbf{m}_1)' \mathbf{D}_1^{-1} (\mathbf{X} - \mathbf{m}_1) \quad (15)$$

$$\begin{aligned} \chi_1^2 &= [0 \quad 1] \times \begin{bmatrix} 3 & -2 \\ -2 & \frac{3}{2} \end{bmatrix} \times \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \frac{3}{2}. \quad (16) \end{aligned}$$

The probability density value is then

$$\Phi_1(\mathbf{X}) = \frac{1}{2\pi} \times \frac{1}{\sqrt{2}} \times e^{-\frac{1}{2} \times \frac{3}{2}} = 0.0532. \quad (17)$$

Similarly, for the second class,

$$\begin{aligned} \chi_2^2 &= [-1 \quad 0] \times \begin{bmatrix} \frac{7}{3} & -\frac{5}{3} \\ -\frac{5}{3} & \frac{4}{3} \end{bmatrix} \times \begin{bmatrix} -1 \\ 0 \end{bmatrix} \\ &= 2 \quad (18) \end{aligned}$$

$$\Phi_2(\mathbf{X}) = \frac{1}{2\pi} \times \frac{1}{\sqrt{3}} \times e^{-\frac{1}{2} \times 2} = 0.0338 \quad (19)$$

Thus, the measurement vector $(4, 3)$ has a higher probability associated with membership in class ω_1 than with membership

in class ω_2 , and it would be appropriate to classify the measurement into ω_1 .

This same decision can be made by using the somewhat simpler decision rule R_1 [expression (4)]:

$$F_{1,1}(\mathbf{X}) = \ln|\mathbf{D}_1| + (\mathbf{X} - \mathbf{m}_1)' \mathbf{D}_1^{-1} (\mathbf{X} - \mathbf{m}_1), \quad (20)$$

$$= \ln(2) + \frac{3}{2} = 2.193; \quad (21)$$

$$F_{1,2}(\mathbf{X}) = \ln|\mathbf{D}_2| + (\mathbf{X} - \mathbf{m}_2)' \mathbf{D}_2^{-1} (\mathbf{X} - \mathbf{m}_2), \quad (22)$$

$$= \ln(3) + 2 = 3.098. \quad (23)$$

Here again the decision is made to classify the observation \mathbf{X} into ω_1 .

The foregoing calculations assume equal probability of membership in ω_1 and ω_2 . Removing this restriction, we take prior probabilities into account. Assume that the following prior probabilities are observed:

$$P\{\omega_1\} = \frac{1}{3}, \quad P\{\omega_2\} = \frac{2}{3}. \quad (24)$$

Recalling the notation from expression (11) that $\Phi_k^*(\mathbf{X}_i)$ denotes the probability density function adjusted for the prior probability,

$$\Phi_k^*(\mathbf{X}_i) = \Phi_k(\mathbf{X}_i) P\{\omega_k\}, \quad (25)$$

we calculate for the two classes

$$\begin{aligned} \Phi_1^*(\mathbf{X}) &= \Phi_1(\mathbf{X}) P\{\omega_1\} = (0.0532) \times \frac{1}{3} \\ &= 0.0177, \end{aligned} \quad (26)$$

$$\begin{aligned} \Phi_2^*(\mathbf{X}) &= \Phi_2(\mathbf{X}) P\{\omega_2\} = (0.0338) \times \frac{2}{3} \\ &= 0.0225. \end{aligned} \quad (27)$$

The actual conditional probabilities [ex-

pression (11)] then become

$$\begin{aligned} P\{\omega_1 | \mathbf{X}_i = (4, 2)\} &= \frac{0.0177}{0.0177 + 0.0225} \\ &= 0.440, \end{aligned} \quad (28)$$

$$\begin{aligned} P\{\omega_2 | \mathbf{X}_i = (4, 2)\} &= \frac{0.0225}{0.0177 + 0.0225} \\ &= 0.560. \end{aligned} \quad (29)$$

Thus, class ω_2 is favored for the observation (4, 3) over class ω_1 .

In terms of the decision rule R_2 , we calculate

$$\begin{aligned} F_{2,1}(\mathbf{X}) &= \ln|\mathbf{D}_1| \\ &\quad + (\mathbf{X} - \mathbf{m}_1)' \mathbf{D}_1^{-1} (\mathbf{X} - \mathbf{m}_1) \\ &\quad - 2 \ln P\{\omega_1\}, \end{aligned} \quad (30)$$

$$= F_{1,1}(\mathbf{X}) - 2 \ln P\{\omega_1\}, \quad (31)$$

$$= 2.193 - 2 \ln\left(\frac{1}{3}\right) = 4.390. \quad (32)$$

For the second class, the outcome is

$$F_{2,2}(\mathbf{X}) = 3.098 - 2 \ln\left(\frac{2}{3}\right) = 3.908. \quad (33)$$

Since the observation is classified into the class which minimizes the value of R_2 , once again the second class is chosen.

Prior Probabilities Contingent on a Single External Conditioning Variable

Having shown how to modify the decision rule to take into account a set of prior probabilities, it is only a small step to consider several sets of probabilities, in which an external information source identifies which set is to be used in the decision rule. As an example, consider the effect of soil type on the distribution of crops that are likely to be grown in an

area. In such a case, a single suite of crops will characterize the entire area, but the expected distribution of crops from one soil type to the next could be expected to vary considerably. Under these circumstances, it would be possible to collect a stratified random sample of the area to be classified, in order to quantify two sets of prior probabilities: one for the crops on the first soil type, the other for crops on the second.

Thus, we introduce a third variable v_i , which indicates the state of the external conditioning variable (e.g., soil type) associated with the observation. We wish, then, to find an expression describing

$$P\{\omega_k | \mathbf{X}_i, v_i\}, \quad (34)$$

the probability that an observation will be a member of the class ω_k , given its vector of observed measurements and the fact that it belongs to class v_i of the external conditioning variable.

In deriving an expression to find this probability, we can make the assumption that the mean vector and dispersion matrix of the class will be the same regardless of the state of the external conditioning variable. This assumption is discussed more fully in a later section. The assumption implies that

$$P\{\mathbf{X}_i | \omega_k\} = P\{\mathbf{X}_i | \omega_k, v_i\}. \quad (35)$$

Expanding both sides of this relationship using the Law of Conditional Probability,

$$\frac{P\{\mathbf{X}_i, \omega_k\}}{P\{\omega_k\}} = \frac{P\{\mathbf{X}_i, \omega_k, v_i\}}{P\{\omega_k, v_i\}}. \quad (36)$$

Solving for the three-way joint probabil-

ity,

$$P\{X_i, \omega_k, v_i\} = \frac{P\{X_i, \omega_k\}P\{\omega_k, v_i\}}{P\{\omega_k\}}. \quad (37)$$

Substituting expression (8) into the left-hand term of the right denominator,

$$P\{X_i, \omega_k, v_i\} = \frac{\Phi_k(X_i)P\{\omega_k\}P\{\omega_k, v_i\}}{P\{\omega_k\}}, \quad (38)$$

$$\begin{aligned} &= \Phi_k(X_i)P\{\omega_k, v_i\} \\ &= \Phi_k^{**}(X_i). \end{aligned} \quad (39)$$

Expanding expression (34) according to the Law of Conditional Probability,

$$P\{\omega_k|X_i, v_i\} = \frac{P\{\omega_k, X_i, v_i\}}{P\{X_i, v_i\}}. \quad (40)$$

Noting that since all classes are included, expression (40), when summed over all classes, must equal 1,

$$\sum_{k=1}^K P\{\omega_k|X_i, v_i\} = 1 = \frac{\sum_{k=1}^K P\{\omega_k, X_i, v_i\}}{P\{X_i, v_i\}}. \quad (41)$$

Rearranging,

$$P\{X_i, v_i\} = \sum_{k=1}^K P\{\omega_k, X_i, v_i\}. \quad (42)$$

Substituting expressions (39) and (42) into

(40), we have

$$\begin{aligned} P\{\omega_k|X_i, v_i\} &= \frac{\Phi_k(X_i)P\{\omega_k, v_i\}}{\sum_{k=1}^K \Phi_k(X_i)P\{\omega_k, v_i\}} \\ &= \frac{\Phi_k^{**}(X_i)}{\sum_{k=1}^K \Phi_k^{**}(X_i)}. \end{aligned} \quad (43)$$

This result is analogous to expression (11); note that the denominator remains constant for all k , and need not actually be calculated to select the class ω_k for which $\Phi_k^{**}(X_i)$ is a maximum.

The application of this expression in classification requires that the joint probabilities $P\{\omega_k, v_i\}$ be known. However, a simpler form using conditional probabilities directly obtained from a stratified random sample can be obtained through the application of the Law of Conditional Probability:

$$P\{\omega_k, v_i\} = P\{\omega_k|v_i\}P\{v_i\}. \quad (44)$$

Since $P\{v_i\}$ cancels from the numerator and denominator after substitution, we have

$$P\{\omega_k|X_i, v_i\} = \frac{\Phi_k(X_i)P\{\omega_k|v_i\}}{\sum_{k=1}^K \Phi_k(X_i)P\{\omega_k|v_i\}}. \quad (45)$$

Thus, either the joint or conditional probabilities may be used in the decision rule:

R_3 : Choose k which minimizes

$$\begin{aligned} F_{3,k}(X_i) &= \ln|\mathbf{D}_k| \\ &\quad + (\mathbf{X}_i - \mathbf{m}_k)' \mathbf{D}_k^{-1} (\mathbf{X}_i - \mathbf{m}_k) \\ &\quad - 2 \ln P\{\omega_k, v_i\} \end{aligned} \quad (46)$$

R'_3 : Choose k which minimizes

$$\begin{aligned} F'_{3,k}(X_i) = & \ln |D_k| \\ & + (X_i - m_k)' D_k^{-1} (X_i - m_k) \\ & - 2 \ln P\{\omega_k | v_i\}. \quad (47) \end{aligned}$$

Numerical example

To illustrate this use of prior probabilities contingent on an external conditioning variable, let us return to the two-class example discussed earlier. This time, however, let us assume that a stratified random sample of the area to be classified produces the estimates of probabilities shown in Table 2. The conditioning variable v_i has two states: v_1 and v_2 . Under the conditions of v_1 , both classes have equal prior probabilities; under v_2 , the second class is more likely to appear, with the probability of 0.7 for ω_2 and 0.3 for ω_1 . Table 3 presents the calculations for this example. For v_1 , ω_1 would be the most likely choice. In the case of v_2 , ω_2 is more probable.

TABLE 2 Simple Prior Probabilities for Numerical Example

	CONDITIONING VARIABLE	
PROBABILITY	v_1	v_2
$P(\omega_1)$	0.5	0.3
$P(\omega_2)$	0.5	0.7

Adding Additional Conditioning Variables

Logic analogous to that of the preceding section shows that classification decisions may be made contingent on any number of external multistate conditioning variables. However, a separate set of prior probabilities must be estimated for all possible states of conditioning variables. For example, consider classifying natural vegetation in an area containing four distinctive rock types, six different soil types, and four unique topographic habitats. Ninety-six sets of prior probabilities will then be required. Estimating these probabilities by separate samples would be prohibitive for such a large number of combinations.

To alleviate this problem, it is possible to model these probabilities from a much smaller set under the assumption of no high-level interaction. This procedure amounts to the calculation of expected values for a multidimensional contingency table when only certain marginal totals are known. Techniques for such modeling have been described in the recent statistical literature, and are summarized in two current books by Bishop et al., (1975) and by Upton (1978). [Other treatments appear in Cox (1970) and Fienburg (1977).] The discussion below is based partly on the treatment presented

TABLE 3 Calculation of Maximum Likelihood Posterior Probabilities

	v_1	v_2	v_1	v_2
	ω_1	ω_2	ω_1	ω_2
$\Phi_k(X_i)$	0.0532	0.0338	0.0532	0.0338
$P(\omega_k v_i)$	0.0266	0.0169	0.0160	0.0237
$\Phi_k(X_i)P(\omega_k v_i)$	0.0266	0.0169	0.0169	0.0237
$\sum_{k=1}^2 \Phi_k(X_i)P(\omega_k v_i)$		0.0435		0.0397
$P(\omega_k X_i, v_i)$	0.611	0.389	0.403	0.597

in pp. 57-101 of Bishop et al., and the reader is referred to these works for cases involving modeling beyond the trivariate case presented here.

As a simple example, consider the three-way case in which a measurement is a member of class ω_k and is also associated with two conditioning variables v_i and o_l . Then the Law of Conditional Probability states

$$P\{\omega_k | v_i, o_l\} = \frac{P\{\omega_k, v_i, o_l\}}{\sum_{k=1}^K P\{\omega_k, v_i, o_l\}}, \quad (48)$$

since

$$P\{v_i, o_l\} = \sum_{k=1}^K P\{\omega_k, v_i, o_l\}.$$

There are $K \times J \times L$ probabilities of the form $P\{\omega_k, v_i, o_l\}$, and we wish to estimate these with maximum likelihood without sampling the full set. Such an estimate is possible, assuming no three-way interaction between ω , v , and o , if probabilities of the forms $P\{\omega_k, v_i\}$, $P\{\omega_k, o_l\}$, and $P\{v_i, o_l\}$ are known.

The method, first described by Deming and Stephan (1940), requires iterative fitting of three-way probabilities $P\{\omega_k, v_i, o_l\}$ to conform with observed two-way probabilities $P\{\omega_k, v_i\}$, $P\{\omega_k, o_l\}$ and $P\{v_i, o_l\}$. Beginning with an initial starting probability $\hat{P}_0\{\omega_k, v_i, o_l\}$, the individual three-way probabilities are first rescaled to conform with one set of two-way probabilities,

$$\begin{aligned} \hat{P}_1\{\omega_k, v_i, o_l\} &= \hat{P}_0\{\omega_k, v_i, o_l\} \\ &\times \frac{\sum_{k=1}^K \hat{P}_0\{\omega_k, v_i, o_l\}}{P\{v_i, o_l\}}. \end{aligned} \quad (49)$$

Rescaling then proceeds for another set of two-way conditionals,

$$\begin{aligned} \hat{P}_2\{\omega_k, v_i, o_l\} &= \hat{P}_1\{\omega_k, v_i, o_l\} \\ &\times \frac{\sum_{j=1}^J \hat{P}_1\{\omega_k, v_j, o_l\}}{P\{\omega_k, o_l\}}, \end{aligned} \quad (50)$$

and finally for the last set,

$$\begin{aligned} \hat{P}_3\{\omega_k, v_i, o_l\} &= \hat{P}_2\{\omega_k, v_i, o_l\} \\ &\times \frac{\sum_{l=1}^L \hat{P}_2\{\omega_k, v_i, o_l\}}{P\{\omega_k, v_i\}}. \end{aligned} \quad (51)$$

As the procedure is repeated, convergence occurs rapidly, and values stabilize within a small number of iterations (Deming and Stephan, 1940). The method always converges toward the unique set of maximum likelihood estimates and can be used with any set of starting values; further, estimates may be determined to any preset level of accuracy (Bishop et al., 1975, p. 83).

In a typical remote sensing application, a stratified random sample is collected which estimates the two conditional probability sets $P\{\omega_k | v_i\}$ and $P\{\omega_k | o_l\}$. In addition, probabilities of the form $P\{\omega_k, o_l\}$ are obtained by processing registered digital images of maps showing the spatial distributions of v and o . By noting that

$$P\{v_i\} = \sum_{l=1}^L P\{v_i, o_l\} \quad (52)$$

and using the Law of Conditional

TABLE 4 Conditional Probabilities for Numerical Example

ω	$P(\omega_k, v_i)$			$P(\omega_k, o_l)$		
	v_1	v_3	v_3	o_1	o_2	o_3
ω_1	0.6	0.5	0.2	0.5	0.8	0.2
ω_2	0.3	0.2	0.4	0.4	0.1	0.3
ω_3	0.1	0.3	0.4	0.1	0.1	0.5

Probability,

$$P\{\omega_k, v_i\} = P\{\omega_k | v_i\} P\{v_i\}, \quad (53)$$

the joint probabilities $P\{\omega_k, v_i\}$ and $P\{\omega_k, o_l\}$ can easily be calculated from the sets of conditional probabilities $P\{\omega_k | v_i\}$ and $P\{\omega_k | o_l\}$.

Numerical example

A simple numerical example will illustrate the iterative method. Table 4 presents a set of one-way conditional probabilities for an example of three classes ω_k , $k=1, 2, 3$ with two conditioning variables v_i , $i=1, 2, 3$, and o_l , $l=1, 2, 3$. Although simple decimal values are assumed here for ease of computa-

tion, these values would normally be obtained by prior random stratified sampling. Also required are the joint probabilities $P\{v_i, o_l\}$ (Table 5). Tables 6 and 7 show how values for $P\{\omega_k, v_i\}$ and $P\{\omega_k, o_l\}$ are calculated according to expressions (52) and (53).

Table 8 presents the results of the first two iterations in fitting the no two-way interaction models to these data. Using the criterion of no further change in any $P\{\omega_k, v_i, o_l\}$ of greater than 10^{-6} , convergence is reached at iteration 23. Although these probabilities can be used directly in decision rule R_3 , it may be easier to examine the values as conditional probabilities as used in R'_3 . These values are shown in the last column of the table.

TABLE 5 Joint Probabilities for Numerical Example

v	$P(v_i, o_l)$			$P(v_i)$
v	o_1	o_2	o_3	
v_1	0.08	0.12	0.14	0.34
v_2	0.07	0.09	0.12	0.29
v_3	0.16	0.10	0.12	0.38
$P(o_l)$	0.31	0.31	0.38	

TABLE 6 Calculation of Joint Two-Way Probabilities for Numerical Example

ω_k	$P(\omega_k v_1)$	$P(v_1)$	$P(\omega_k, v_1)$	$P(\omega_k v_2)$	$P(v_2)$	$P(\omega_k, v_2)$	$P(\omega_k v_3)$	$P(v_3)$	$P(\omega_k, v_3)$						
ω_1	0.6	x	0.34	-	0.204	0.5	x	0.28	-	0.140	0.2	x	0.38	-	0.076
ω_2	0.3	x	0.34	-	0.102	0.2	x	0.28	-	0.056	0.4	x	0.38	-	0.152
ω_3	0.1	x	0.34	-	0.034	0.3	x	0.28	-	0.084	0.4	x	0.38	-	0.152

TABLE 7 Calculation of Joint Two-Way Probabilities for Numerical Example

ω_k	$P(\omega_k o_1)$	$P(o_1)$	$P(\omega_k, o_1)$	$P(\omega_k o_2)$	$P(o_2)$	$P(\omega_k, o_2)$	$P(\omega_k o_3)$	$P(o_3)$	$P(\omega_k, o_3)$						
ω_1	0.5	x	0.31	-	0.155	0.8	x	0.31	-	0.248	0.2	x	0.38	-	0.076
ω_2	0.4	x	0.31	-	0.124	0.1	x	0.31	-	0.031	0.3	x	0.38	-	0.114
ω_3	0.1	x	0.31	-	0.031	0.1	x	0.31	-	0.031	0.5	x	0.38	-	0.190

**ORIGINAL PAGE IS
OF POOR QUALITY**

USE OF PRIOR PROBABILITIES

147

TABLE 8 Iterative Fitting of No-Three-Way Interaction Model

k	l	INITIAL		ITERATION 1				ITERATION 2				FINAL	
		$\hat{P}_0(\omega_k, r_j, \sigma_l)$	$\hat{P}_1(\omega_k, r_j, \sigma_l)$	$\hat{P}_2(\omega_k, r_j, \sigma_l)$	$\hat{P}_3(\omega_k, r_j, \sigma_l)$	$\hat{P}_1(\omega_k, r_j, \sigma_l)$	$\hat{P}_2(\omega_k, r_j, \sigma_l)$	$\hat{P}_3(\omega_k, r_j, \sigma_l)$	$\hat{P}_1(\omega_k, r_j, \sigma_l)$	$\hat{P}_2(\omega_k, r_j, \sigma_l)$	$\hat{P}_3(\omega_k, r_j, \sigma_l)$	$\hat{P}(\omega_k r_j, \sigma_l)$	$\hat{P}(\omega_k r_j, \sigma_l)$
1	1 1	0.0370	0.0680	0	.53	0.0502	0.0487	0.0635	0.0557	0.0614	0.0614	0.7674	
1	1 2	0.0370	0.0680	0.1205	0.1074	0.1043	0.1198	0.1098	0.1132	0.1132	0.9431		
1	1 3	0.0370	0.0680	0.0369	0.0525	0.0510	0.0457	0.0494	0.0456	0.0456	0.3546		
1	2 1	0.0370	0.0467	0.0517	0.0432	0.0408	0.0532	0.0479	0.0517	0.0517	0.7383		
1	2 2	0.0370	0.0467	0.0426	0.0760	0.0718	0.0823	0.0785	0.0820	0.0820	0.9116		
1	2 3	0.0370	0.0467	0.0253	0.0263	0.0274	0.0245	0.0250	0.0221	0.0221	0.1845		
1	3 1	0.0370	0.0253	0.0280	0.0422	0.0293	0.0382	0.0424	0.0386	0.0386	0.2413		
1	3 2	0.0370	0.0253	0.0449	0.0579	0.0402	0.0461	0.0542	0.0504	0.0504	0.5038		
1	3 3	0.0370	0.0253	0.0138	0.0093	0.0065	0.0058	0.0052	0.0052	0.0023	0.0196		
2	1 1	0.0370	0.0340	0.0408	0.0272	0.0309	0.0262	0.0230	0.0179	0.0179	0.2243		
2	1 2	0.0370	0.0340	0.0102	0.0091	0.0103	0.0098	0.0091	0.0059	0.0059	0.0492		
2	1 3	0.0370	0.0340	0.0375	0.0534	0.0607	0.0544	0.0589	0.0613	0.0613	0.4376		
2	2 1	0.0370	0.0187	0.0224	0.0187	0.0221	0.0187	0.0169	0.0153	0.0153	0.2187		
2	2 2	0.0370	0.0187	0.0056	0.0051	0.0061	0.0052	0.0049	0.0043	0.0043	0.0482		
2	2 3	0.0370	0.0187	0.0206	0.0235	0.0278	0.0249	0.0254	0.0277	0.0277	0.2307		
2	3 1	0.0370	0.0507	0.0608	0.0615	0.0633	0.0790	0.0897	0.0931	0.0931	0.5817		
2	3 2	0.0370	0.0507	0.0152	0.0198	0.0200	0.0170	0.0200	0.0217	0.0217	0.2169		
2	3 3	0.0370	0.0507	0.0599	0.0380	0.0387	0.0347	0.0313	0.0239	0.0239	0.1983		
3	1 1	0.0370	0.0113	0.0039	0.0026	0.0022	0.0016	0.0014	0.0007	0.0007	0.0083		
3	1 2	0.0370	0.0113	0.0039	0.0035	0.0029	0.0023	0.0021	0.0009	0.0009	0.0077		
3	1 3	0.0370	0.0113	0.0239	0.0341	0.0288	0.0283	0.0317	0.0291	0.0291	0.2079		
3	2 1	0.0370	0.0280	0.0096	0.0081	0.0090	0.0058	0.0052	0.0030	0.0030	0.0430		
3	2 2	0.0370	0.0280	0.0096	0.0069	0.0068	0.0068	0.0065	0.0065	0.0065	0.0401		
3	2 3	0.0370	0.0280	0.0591	0.0675	0.0672	0.0683	0.0683	0.0672	0.0672	0.5848		
3	3 1	0.0370	0.0507	0.0175	0.0263	0.0329	0.0236	0.0288	0.0283	0.0283	0.1770		
3	3 2	0.0370	0.0507	0.0175	0.0225	0.0282	0.0219	0.0257	0.0279	0.0279	0.2793		
3	3 3	0.0370	0.0507	0.1070	0.0727	0.0910	0.0924	0.0937	0.0937	0.0937	0.7811		

the outcome of a classification contingent on the earlier classification. This application is best clarified by an example. Consider an agricultural classification with four field types: rice, cotton, orchard, and fallow. An early spring classification reveals the presence of young rice with high accuracy, but at that time cotton cannot be distinguished from fallow fields. Orchards are easily distinguished from field crops at any time of year. By early summer, many fields which classified as fallow are likely to be in cotton; however, fields classified as rice are still likely to be rice. Orchards will remain unchanged in areal extent.

Data from prior years are collected to quantify these expected changes, and a transition probability matrix is devised which describes the changes in classification expected during the early spring-early summer period (Table 9). In this example, spring classification shows 30% of the observations to be rice; by summer, 90% of these observations are expected to continue as rice, with 10% returning to fallow because of crop failure or lack of irrigation. Twenty percent of the spring observations are orchards, and all of these are expected to remain in orchard through early summer. Fallow fields, constituting 50% observations, are most likely to become cotton (probability = 0.7), with a few becoming rice, orchard,

or remaining fallow (probabilities = 0.1). Since no observations are classified as cotton in spring, no transition probabilities are needed for that class. This use of transition probabilities was suggested as early as 1967 by Simonett et al.

The transition probability matrix can also be recognized as a matrix of conditional probabilities $P(\omega_k | \nu_i)$ which describe the probability that an observation will fall into summer class ω_k given that the observation falls into spring class ν_i . Thus, the early summer classification can be made contingent on the early spring classification through the prior probability mechanisms discussed earlier, and any possible confusion between cotton and rice in summer will be resolved by the spring classification. It is also interesting to note that the transition probability matrix is actually a square stochastic matrix, and therefore the situation is equivalent to a simple one-step Markov process. Under these conditions, the expected posterior probabilities $P(\omega_k)$ are

$$P(\omega_k) = \sum_{i=1}^J P(\omega_k | \nu_i) P(\nu_i). \quad (54)$$

In a recent paper, Swain (1978) has carried this approach a step further, incorporating in the decision rule both

TABLE 9 Agricultural Time-Sequential Classification Example

SPRING CLASS	P(ν_i)	P($\omega_k \nu_i$)			
		ω_1	ω_2	ω_3	ω_4
		RICE	COTTON	ORCHARD	FALLOW
ν_1	Rice	0.3	0.9	0.0	0.1
ν_2	Cotton	0.0	—	—	—
ν_3	Orchard	0.2	0.0	0.0	0.0
ν_4	Fallow	0.5	0.1	0.7	0.2
	$P(\omega_k)$	0.32	0.35	0.25	0.06

measurement vectors $\mathbf{X}_{t,1}$ and $\mathbf{X}_{t,2}$ taken from times t_1 and t_2 at point i in space. In contrast, the approach described above uses $\mathbf{X}_{t,1}$ to predict v_i at time t_1 and then uses $\mathbf{X}_{t,2}$ and v_i to make the classification decision at time t_2 . Swain's decision rule, in the notation of this paper, becomes

R_S : Choose k to maximize

$$F_{S,k}(\mathbf{X}_{t,1}, \mathbf{X}_{t,2}) = \sum_{j=1}^J P\{\mathbf{X}_{t,1}|v_j\} P\{\mathbf{X}_{t,2}|\omega_k\} P\{\omega_k|v_j\} P\{v_j\}. \quad (55)$$

Swain has termed this rule the "cascade classifier."

Swain's approach has the advantage of using full information about the distances of the measurement vectors $\mathbf{X}_{t,1}$ and $\mathbf{X}_{t,2}$ from class means; however, as Swain notes, there is no way to make the first observation set dominate the second. When the transition probability matrix goes to an identity matrix, the classification rule becomes:

$$F_{S,k}(\mathbf{X}_{t,1}, \mathbf{X}_{t,2}) = P\{\mathbf{X}_{t,1}|v_j\} P\{\mathbf{X}_{t,2}|\omega_k\} P\{v_j\}, \quad (56)$$

and the two observations becomes equally weighted. Decision rule R_3 does allow the first observation to dominate; here, an identity transition probability matrix will preserve the first classification completely. On the other hand, R_3 assumes that the prior classification is perfectly correct, and any errors in the prior classification will also be preserved to an extent controlled by the transition probabilities. Thus, both approaches are relevant,

depending on the classification task at hand.

Remote Sensing Example

The preceding numerical examples have demonstrated the application of prior probabilities in maximum likelihood classification in a computational context; a real example drawn from remote sensing should serve to further enhance understanding in an operational context. This example (Strahler et al., 1978) is drawn from a problem involving classification of natural vegetation in a heavily forested area of northern California. In the classification, spectral data are used to define species-specific timber types, and elevation and slope aspect are used as collateral data channels to improve classification accuracy.

The area selected for application of the classification techniques described above is referred to as the Doggett Creek study area, comprising about 220 km² of private and publicly-owned forest land in northern California near the town of Klamath River. Located within the Siskiyou Mountains, elevations in the area range from 500 m at the Klamath River, which crosses the southern portion, to 2065 m near Dry Lake Lookout on an unnamed summit. A well developed network of logging roads and trails is present, providing relatively easy access to nearly all of the area by road or foot.

A wide variety of distinctive vegetation types is present in the area. Life-form classes include alpine meadow, fir park, pasture, cropland, and burned reforested areas. Forest vegetation includes, from high elevation to low elevation, such types as red fir, white fir, douglas fir-ponderosa pine-incense cedar, pine-oak,

and oak-chapparal. Thus, the topographic and vegetational characteristics of the area are well differentiated.

After a review of available Landsat frames which included the Doggett Creek area, two were selected for analysis: 4 July 1973, and 15 October 1974. The two frames were obtained as computer compatible tapes from the EROS Data Center, Sioux Falls, SD and then reformatted and precision rectified to sinusoidal projections. Pixel size was converted to 80×80 m in the rectification process to facilitate film-writer playback. Using the July image as a base, the October frame was registered to within a half-pixel error using seven control points. In this process, the October image was resampled to conform with the July image using a cubic spline convolution algorithm. Figure 1 presents an image of the study area using Landsat band 5 (0.6–0.7 μm) from the July frame.

Also registered to the July image was a terrain image derived from the U.S. Geological Survey 1:250,000 digital terrain tape for the Weed, CA, quadrangle. In the registration process, the image was converted to 80×80 m pixel size, and stretched to yield a full range of gray-tone values. Slope and aspect images were generated directly from the registered elevation data using a least-squares algorithm which fits a plane to each pixel and its four nearest neighbors.

The slope aspect image consisted initially of gray-tone densities between 0 (black) and 255 (white) which indicated the azimuth of slope orientation, ranging clockwise from 0° to 359° . These values were then transformed by a cosine function proposed by Hartung and Lloyd (1969). Since northeast slopes present the most favorable growing environment, and

southwest slopes the least favorable, with northwest and southeast slopes of neutral character, the density tones of azimuths were rescaled with 3 representing due southwest and 255 representing due northeast. (Values of 0, 1, and 2 were reserved for special codings.) Neutral slopes, oriented northwest or southeast, thus received density tones near 127. The function also corrected automatically for the 12° skew of the Landsat image. For processing as collateral data channels, both elevation and transformed aspect were converted to three-state variables: elevation to low, middle, and high; and aspect to southwest, neutral (southeast or northwest) and northeast. Figure 2 shows elevation and aspect images as well as their three-state versions.

Following an initial reconnaissance of the area, thirteen species-specific forest cover classes were selected as representing the range of cover types within the study area. These classes were defined by a set of 93 training sites ranging in size from approximately twenty to one hundred pixels. Further processing revealed the presence of several subtypes within most of the forest cover classes. For example, open-canopy douglas fir training sites were divided into two subtypes. Such subtypes were also defined for hardwood, white fir, douglas fir, sparse, and grass and shrub cover classes. Throughout the classification procedure, these subtypes were kept separate, joining together only in the final classification map.

In order to obtain estimates of prior probabilities for the forest cover class types, one hundred points were randomly selected from a grid covering the Doggett Creek study area by drawing coordinates from a random number table. At

ORIGINAL PAGE IS
OF POOR QUALITY.

USE OF PRIOR PROBABILITIES

151

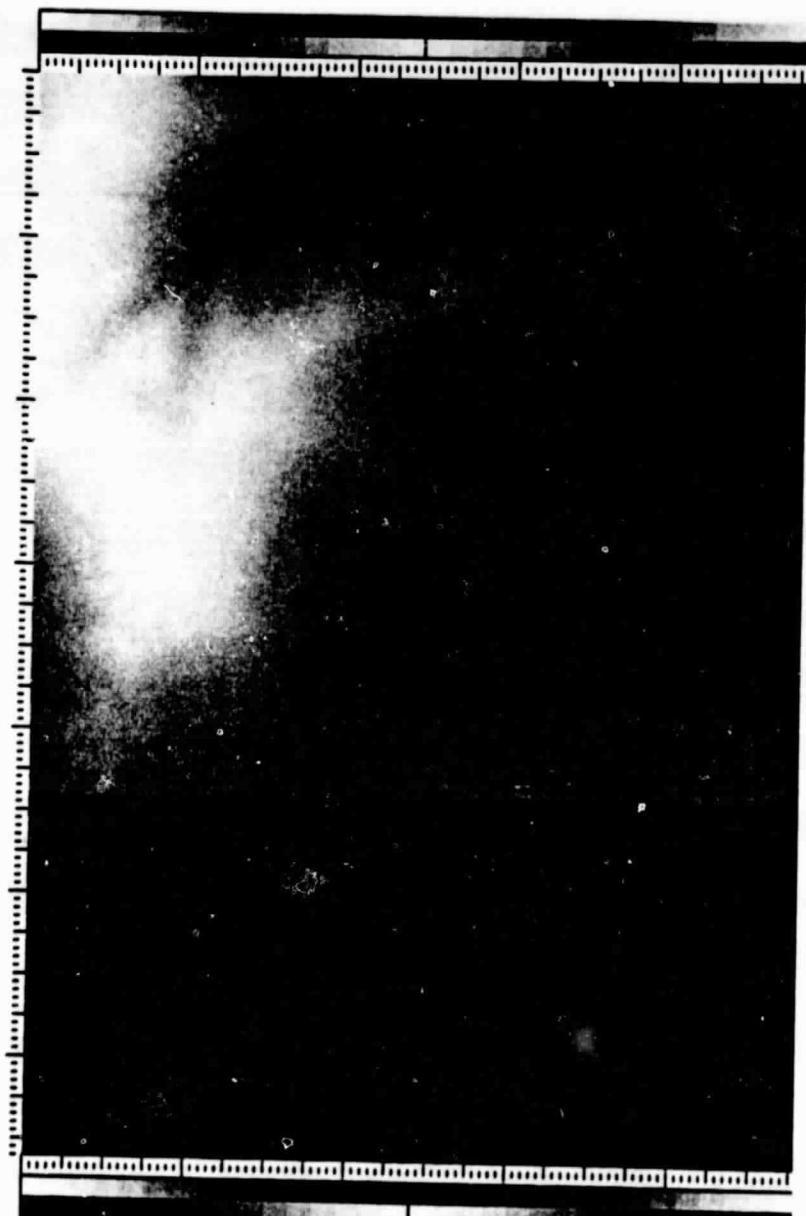


FIGURE 1. Landsat band 5 image of Doggett Creek study area, Klamath National Forest, CA

ORIGINAL PAGE IS
OF POOR QUALITY

152

ALAN H. STRAHLER



(a)

FIGURE 2. Registered continuous and trilevel elevation (a, b) and aspect (c, d) images of Doggett Creek study area.

ORIGINAL PAGE IS
OF POOR QUALITY

153

USE OF PRIOR PROBABILITIES



(b)

FIGURE 2.

ORIGINAL PAGE IS
OF POOR QUALITY

154

ALAN H. STRAHLER



DOGGETT CREEK, CALIFORNIA
COMPASS ASPECT
LOW PASS FILTER OF FUNCTION IMAGE

DOGASPEX SAT SEP 3, 1977 114811 JPL/IPL
(c)

FIGURE 2.

ORIGINAL PAGE IS
OF POOR QUALITY

155

USE OF PRIOR PROBABILITIES



(d)

FIGURE 2.

ORIGINAL PAGE IS
OF POOR QUALITY

156

ALAN H. STRAHLER

each of these points, the forest cover class was determined either by interpretation of 1:15,840 color aerial photography, or by actual field visit. Of the 100 points, 15 were discarded because they fell (1) in locations which were inaccessible in the time available; or (2) outside the area covered by 1:15,840 air photos (and therefore could not be accurately located on either the Landsat frame or on the ground). At each point, the elevation and aspect class was also recorded, thus allowing type counts to be cross tabulated according to elevation and aspect.

From this sample of 85 points, three sets of probabilities were prepared. The first of these recorded the unconditional prior probabilities of the forest cover types—that is, their proportional representation within the entire study area. The second and third sets of probabilities aggregated the points according to elevation and aspect classes, and were used to estimate three sets of probabilities for each topographic parameter (low, middle, and high for elevation, and northeast, neutral and southwest for aspect). Table 10 shows how the classes were defined, and describes the number of points falling into each.

These estimates of probabilities lack precision because the number of sample

points is small; with 85 samples distributed across 13 cover types, the calculated probabilities are more likely to indicate adequately the rank order of the magnitudes rather than the true values of the magnitudes themselves. However, under constraints of field time and expense, it was not possible to prepare a larger dataset for this particular trial. Considering the sensitivity of the classification to extreme probability values, future work should estimate these probability sets to a higher degree of accuracy.

This dataset was also used to estimate classification accuracies. By recording the pixel location of each of the sample points, the cover type as determined on the ground could be compared with the cover type as classified on the Landsat image. Because of uncertainties in locating each pixel on the 1:15,840 air photos, it was necessary to specify alternate acceptable classifications for each point. For example, a pixel falling into a stand identified on the ground as douglas fir, open canopy, might fall almost entirely on a clearing, and thus be classified as grass/shrub, or sparse if containing a few canopy trees. In such a case, the classification was termed correct. On the other hand, classifications such as hardwood, alpine meadow, or red fir forest would be an obvious error in a

TABLE 10 Elevation and Aspect Class Definitions

CODE	DEFINITION	POINT COUNT
Elevation		
Low	< 1067 m	45
Middle	1068–1524 m	26
High	> 1525 m	14
Aspect		
Northeast	337.6°–112.5°	26
Neutral	122.6°–157.5°; 292.6°–337.5°	25
Southwest	157.6°–292.5°	34

douglas fir stand. Here again, estimated accuracies are influenced by the limited size of the sample.

Note that the field data are used to produce a classification which is then assessed for accuracy by reference to the same data. Separate samples would clearly be more desirable. The decision to use the same set of samples to determine accuracy that was used to estimate the probability sets was, again, influenced by available field time and travel funds. However, the accuracies are greatly dependent on the spatial location of the data points; only in the aggregate does each data point influence the classification. Thus, we would expect the accuracies to reflect only a slight positive bias produced by this cost-reducing strategy.

Although several different classifications were carried out, only three are of importance here. The first used spectral data only, and assumed equal prior probabilities; this classification yielded an accuracy of 58% (Fig. 3). In the second, three sets of prior probabilities for the forest types were used, each contingent on one of the three elevation states (Table 10; Fig. 4). The classification software was modified to use a table look-up of prior probabilities with elevation class as one index into the prior probability table. This technique increased classification accuracy from 58 to 71%.

The third classification used two sets of prior probabilities contingent on elevation and aspect, analogous to $P\{\omega_k | v_i\}$ and $P\{\omega_k | o_l\}$ in the preceding section. Software then systematically sampled the registered elevation class and terrain class images to yield the joint probabilities of elevation and aspect classes, analogous to $P\{v_i, o_l\}$. The iterative algorithm de-

scribed earlier was then applied to estimate the set of conditional probabilities for forest cover classes contingent on all combinations of elevation and aspect classes. Classification using these estimated probabilities contingent on both elevation and aspect yielded an accuracy of 77%, an improvement over that observed for elevation alone (Fig. 5). Thus, this example demonstrates how prior probabilities can be used to merge continuous variables of multispectral brightness with discrete variables of elevation and aspect class to improve classification accuracies.

Discussion

As noted earlier, extreme values of prior probabilities can force a classification to be made essentially without information concerning the observation itself. When priors are equal, however, they have no effect. The classifier, then, continuously trades off the role of the multivariate information for the role of prior information, depending on both the magnitude of the distance of the multivariate observation vector from the class mean vector and the ratios of the particular prior probabilities involved in the decision. When the experimental design allows the priors to be determined by external conditioning variables, the effect is to classify based on multivariate information when the possible classes are not particularly influenced by the conditioning variable and to classify based on prior information when multivariate data are equivocal or some classes are much more or less likely than others. Thus, in the forest classification example, terrain information served to differentiate species-specific cover types (e.g., red fir, white

ORIGINAL PAGE IS
OF POOR QUALITY

158

ALAN H. STRAHLER

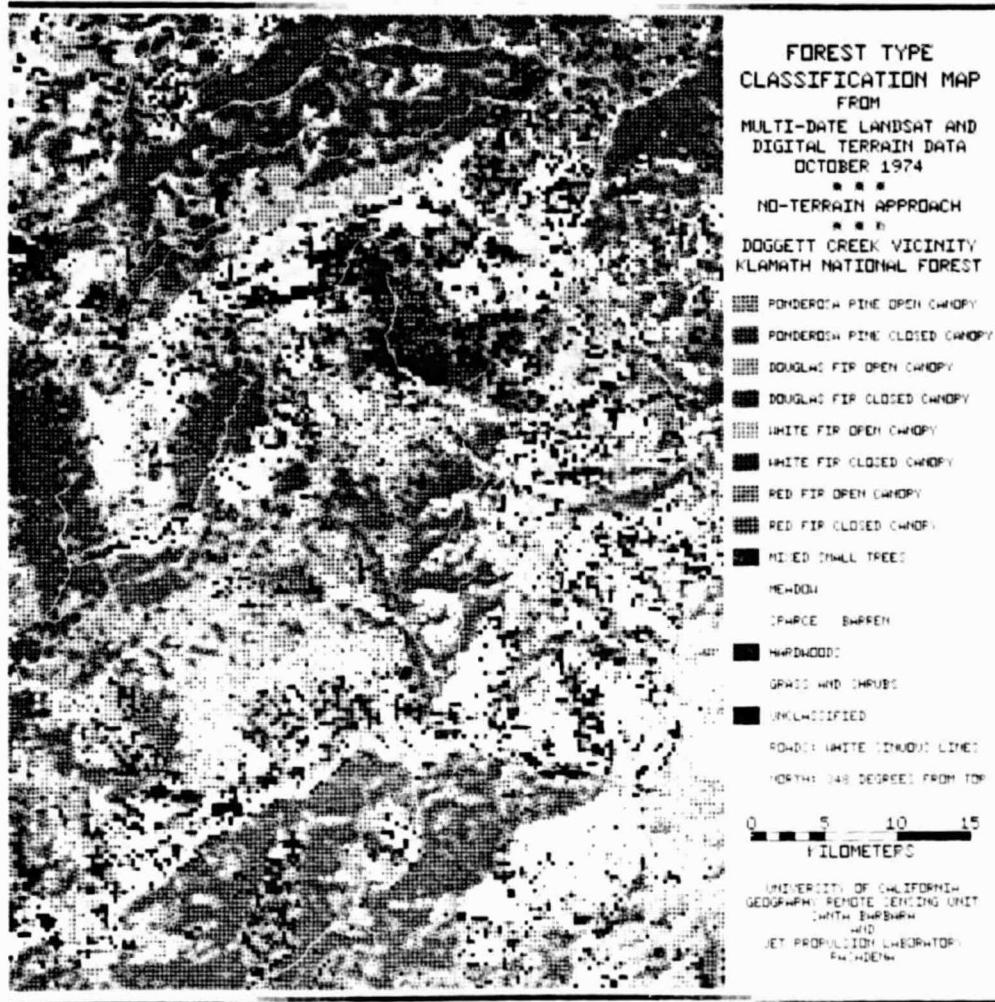


FIGURE 3. Classification map based on spectral data only; accuracy, 58%.

fir), whereas spectral information differentiated life form classes (e.g., grass-shrub, hardwood).

Another important point concerns the dependence of the prior probabilities on the scene itself; the relative areas of the classes in the output scene must be accurately estimated. If the output scene shifts in area, then the priors must be changed. The classification cannot be extended to a new area without reestimating the prior probabilities. Thus, it would be ap-

propriate to use county crop acreage values to set prior probabilities only when the entire area of the county, no more, no less, is to be classified. In the case of one or two external conditioning variables determining the appropriate set of priors, both the joint probabilities $P\{v_i, o_l\}$ and the conditional probabilities $P\{\omega_k | v_i\}$ and $P\{\omega_k | o_l\}$ must truly represent the area to be classified, for, taken together, they determine the prior probability values actually used in computa-

ORIGINAL PAGE IS
OF POOR QUALITY

USE OF PRIOR PROBABILITIES

159



FIGURE 4. Classification map based on spectral data, with elevation included by varying prior probabilities. Key to map symbols is included in Fig. 3. Accuracy is 71%.

ORIGINAL PAGE IS
OF POOR QUALITY

160

ALAN H. STRAHLER



FIGURE 5. Classification map using spectral, elevation, and aspect data. Key to map symbols is included in Fig. 3. Accuracy is 77%.

tion. In some applications, it may be possible to extend the conditionals to a new scene in which only the joint probabilities of the variables change—for example, a forest cover classification with elevation and aspect as conditioning variables which is extended from one uniform area to another. The new area will have different joint probabilities $P\{\nu_i, o_i\}$ (and derived priors $P\{\nu_i\}$ and $P\{o_i\}\}$, but it might be reasonable to assume that the conditional probabilities are ecologically based and remain consistent from one area to the next.

It should be noted that collateral information cannot be incorporated through the prior probability mechanism without the collection of data to estimate the priors and/or conditionals. If the collateral data are likely to be unrelated to the multivariate data and are expected to influence strongly the prior probabilities of the classes, then such estimating costs will be justified, for significant improvements in accuracy should result. Ultimately, it is up to the user to balance the costs of acquiring such information with the value of the expected payoffs in accuracy which are anticipated.

The logic which culminates in decision rules R_2 and R_3 assumes that the mean vector and dispersion matrix for a class are not affected by the external conditioning variable [see expression (35)]—in the remote sensing case, this means that the signatures are invariant. In some applications, this assumption may not be warranted. An example would be an agricultural crop classification with soil type as a collateral variable. Here the signature of the soil itself, at least in the earlier stages of crop development, will influence the crop signature. In this situation, there is no recourse but to

spectrally characterize each combination of crop and soil type so that probabilities of the form $P\{\omega_k | X_i, \nu_i\}$ can be calculated. Following the logic of expressions (5) through (11), it is possible to show that

$$P\{\omega_k | X_i, \nu_i\} = \frac{P\{X_i | \omega_k, \nu_i\} P\{\omega_k | \nu_i\}}{\sum_{k=1}^K P\{X_i | \omega_k, \nu_i\}}, \quad (57)$$

which could be made the basis of a decision rule related to R_2 .

A final point worthy of discussion concerns modeling of joint probabilities, suggested in an earlier section to reduce the need for more extensive ground sampling. The model presented is but one example of a large variety of techniques by which collateral data can be used to predict the spatial distribution of classes in an output image. Discrete and continuous collateral variables can be merged either using empirical techniques including multiple regression, logit analysis, discriminant analysis, analysis of covariance, and contingency table analysis, or by constructing more functional models which model the spatial processes actually occurring in a deterministic way. When such models are constructed and interfaced with remotely sensed data, the result may be extremely powerful, both for the ability to accurately predict a spatial pattern and for the understanding of the complex system which produces it.

Conclusions

The use of prior probabilities in maximum likelihood classification allows:

- (1). the incorporation into the classification of prior knowledge concerning

- the frequencies of output classes which are expected in the area to be classified;
- (2). the merging of one or more discrete collateral datasets into the classification process through the use of multiple prior probability sets describing the expected class distribution for each combination of discrete collateral variables; and
 - (3). the use of time-sequential information in making the outcome of a later classification contingent on an earlier classification.

Thus, prior probabilities can be a powerful and effective aid to improving classification accuracy and modeling the behavior of spatial systems.

The research reported herein was supported in part by NASA grant NSG-2377, NASA contract NAS-9-15509, and the California Institute of Technology's President's Fund (award PF-123), under NASA contract NAS-7-100. I am greatly indebted to D. S. Simonett, P. H. Swain, R. M. Haralick, and W. Wigton for critical review of the manuscript.

References

- Addington, J. D. (1975), *VICAR Program FASTCLAS*, VICAR Documentation, Image Processing Laboratory, Jet Propulsion Laboratory, California Inst. of Technology, Pasadena, CA, 3 pp.
- Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975), *Discrete Multivariate Analysis: Theory and Practice*, MIT Press, Cambridge, MA, 557 pp.
- Brooner, W. G., Haralick, R. M., and Dinstein, I. (1971), Spectral parameters affecting automated image interpretation using Bayesian probability techniques, *Proceedings of the Seventh International Symposium on Remote Sensing of Environ.*, ERIM, Ann Arbor, MI, pp. 1929-1948.
- Chow, C. K. (1957), An optimum character recognition system using decision functions, *IREE Trans. Electron. Computers* 6:247-254.
- Cox, D. R. (1970), *The Analysis of Binary Data*, Methuen & Co., Ltd., London, 142 pp.
- Deming, W. E. and Stephan, F. F. (1940), On a least squares adjustment of a sampled frequency table when the expected marginal totals are known, *Ann. Math. Stat.* 11:427-444.
- Dubes, R. and Jain, A. K. (1976), Clustering techniques: The user's dilemma, *Pattern Recog.* 8:247-260.
- Fienberg, S. E. (1977), *The Analysis of Cross-Classified Categorical Data*, MIT Press, Cambridge, MA, 151 pp.
- Goodenough, D., and Shlien, S. (1974), Results of cover-type classification by maximum likelihood and parallelepiped methods, *Proceedings of the Second Canadian Symposium on Remote Sensing*, 1:136-164. Canada Centre for Remote Sensing, Ottawa.
- Hartung, R. E., and Lloyd, W. J. (1969), Influence of aspect on forests of the Clarksville soils in Dent County, Missouri, *J. Forestry* 67:178-182.
- Kendall, M. G. (1972), in *Frontiers of Pattern Recognition* (Satosi Watanabe, Ed.), Academic Press, New York, pp. 291-307.
- Nilsson, N. J. (1965), *Learning Machines—Foundations of Trainable Pattern-Classifying Systems*, McGraw-Hill Book Co., New York.
- Reeves, R. G., Anson, A., and Landen, D. (1975), *Manual of Remote Sensing*, Amer. Soc. of Photogramm., Falls Church, VA, 2 vols., 2144 pp.
- Richardson, A. J. and Wiegand, C. L. (1977), Distinguishing vegetation from soil background information, *Photogrammetric Engr. Remote Sens.* 42:1541-1552.

- Rouse, J. W., Jr., Hass, R. H., Schell, J. A., and Deering, D. W. (1973), Monitoring vegetation systems in the Great Plains with ERTS, *Third ERTS Symposium*, NASA Special Publication SP-351, 1:309-317.
- Sebestyen, G. (1962), *Decision-Making Processes in Pattern Recognition*, Macmillan, New York.
- Schell, J. A. (1973), in *Remote Sensing of Earth Resources, Volume I* (F. Shahrokhi, ed.), Univ. of Tennessee Space Inst., Tullahoma, TN, pp. 374-394.
- Schlien, S., and Smith, A. (1975), A rapid method to generate spectral theme classification of Landsat imagery, *Remote Sens. Environ.* 4:67-77.
- Simonett, D. S., Eagleman, J. E., Erhart, A. B., Rhodes, D. C., and Schwarz, D. E. (1967), *The Potential of Radar as a Remote Sensor in Agriculture: 1. A Study with K-Band Imagery in Western Kansas*, CRES Report No. 16-21, University of Kansas, Lawrence, KN, 13 pp.
- Strahler, A. H., Logan, T. L., and Bryant, N. A. (1978), Improving forest cover classification accuracy from Landsat by incorporating topographic information, *Proceedings of the Twelfth International Symposium on Remote Sensing of Environ.*, ERIM, Ann Arbor, MI, pp. 927-942.
- Swain, P. H. (1978), Bayesian classification in a time-varying environment, *IEEE Trans. Systems, Man, and Cybern.*, SMC-8:879-883.
- Tatsuoka, M. M. (1971), *Multivariate Analysis: Techniques for Educational and Psychological Research*, John Wiley & Sons, New York, 310 pp.
- Tatsuoka, M. M. and Tideman, D. V. (1954), Discriminant Analysis, *Rev. of Ed. Res.* 25:402-420.
- Upton, G. J. G. (1978), *The Analysis of Cross-Tabulated Data*, John Wiley & Sons, New York, 148 pp.

Received 20 November 1979.

A P P E N D I X B

**ORIGINAL PAGE IS
OF POOR QUALITY**

INCORPORATING COLLATERAL DATA IN LANDSAT

CLASSIFICATION AND MODELING PROCEDURES

A.H. Strahler, J.E. Estes, P.F. Maynard,
F.C. Mertz, D.A. Stow

Department of Geography
University of California
Santa Barbara, CA, 93106, U.S.A.

ABSTRACT

A number of existing statistical techniques can be used to merge spectral image data with collateral information, including regression, ANOVA, MANOVA, ANCOVA, MANCOVA, discriminant analysis, maximum likelihood classification with or without prior probabilities, contingency table analysis, and logit modeling. The choice of an appropriate technique depends upon the nature of input and output variables -- continuous, discrete, or categorical -- and the appropriate model -- parametric or nonparametric.

Logit modeling is a very versatile technique which is well adapted to remote sensing application. The logit, which is the natural logarithm of an odds ratio for two states of an output categorical variable, is predicted by a linear or curvilinear function of continuous or categorical input variables. Since the logit models probabilities or proportions, it can be used directly as a classifier or indirectly as an estimator of prior probabilities for conventional maximum likelihood classification with prior probabilities. The logit model is nonparametric, a feature which makes it highly desirable when used to merge disparate types of collateral data. The disadvantages of the logit model are that more calibration (training) data are required to fit the model, and that fitting requires an iterative nonlinear optimization procedure. A logit model was devised and fitted to forest species compositional data for northern California, predicting the proportion of timber volume in each of five species at each pixel based on registered terrain data quantifying elevation and slope aspect.

Another versatile tool is maximum likelihood classification with prior probabilities. By making prior probabilities conditional on a collateral data channel, the information contained within the channel can be conveyed to the maximum likelihood algorithm. An example is in land use, in which a previous classification and an externally devised transition probability matrix are used together with new image data to produce an updated classification consistent with the observed pattern of change. This technique has relevance for monitoring urban expansion and the impact of forest clearing in developing nations.

1. INTRODUCTION

Viewed in a broad context, the problem of combining image data and spatial collateral data into a predicted output map or image is actually a problem of combining continuous and categorical variables in a modeling framework capable of producing continuous or categorical outputs. At the University of California, Santa Barbara, NASA-supported research (grant NSG-2377) is currently underway to identify existing models and procedures for spatial modeling and to apply them to selected datasets to demonstrate their applicability in a remote sensing situation.

Collateral data, here defined as preexisting spatial information in the form of a map, pro-

**ORIGINAL PAGE IS
OF POOR QUALITY**

cessed image, or set of observations at grid coordinate locations, can be combined with Landsat or other remotely sensed digital imagery to enhance classification accuracy or to construct models which predict spatial patterns of ground phenomena. Collateral information can be either continuous or categorical in nature. Examples are elevation, slope, or aspect channels obtained from a digital terrain model (continuous); and rock type, soil type, crop type, or land use (categorical). Image data, with which collateral information are to be combined, are typically continuous in nature, although values may be quantized into discrete gray tone levels for data processing applications. Desired outputs may also be either categorical or continuous. Any type of classification constitutes a discrete or categorical output, whereas such outputs as percent bare ground, timber volume, soil loss, or forage cover are continuous in character.

Figure 1 presents appropriate techniques for the merging of continuous, categorical and mixed (both continuous and categorical) datasets to produce either continuous or categorical outputs. Continuous output models, including regression, analysis of variance, and analysis of covariance, are all mathematically related and based on least squares algebra. Categorical output models are more diverse and include variance maximizing techniques such as discriminant analysis as well as the nonparametric methods of contingency table analysis and logit modeling. Maximum likelihood classification may be viewed as a special case of logit modeling in which the input variables are assumed to be normally distributed. Nonparametric and maximum likelihood techniques, because they produce probabilities of classification as an output, also have the advantage that they can be adjusted for prior probabilities. Because many present remote sensing applications call for categorical classification, these latter methods are probably most useful in combining continuous image data with categorical and continuous types of collateral data.

Demonstrations of a selected set of these techniques are planned and under current development; their current status is discussed in following sections. Several of these examples have important implications for remote sensing in developing nations. In one application, logit modeling is used to fit a model which describes the probability of occurrence of various forest species given elevation and slope aspect values obtained from a registered digital terrain model. Once obtained, these probabilities can be used as prior probabilities in a maximum likelihood classification of a Landsat image with registered terrain data for natural vegetation units. This technique could facilitate the accurate identification of forest types in complex tropical upland environments.

In another example, land use classification for change detection monitoring is improved by a contingency table-analytic technique which quantifies the probabilities of change for each land use type during the fixed time interval. This technique has important implications for many developing nations, especially in Central America, where urban expansion is impacting agricultural land, and forest clearing for agriculture is impacting large natural environments. Additional examples are being developed, focused on Landsat and collateral datasets obtained for an area of Ventura County, California.

2. STATISTICAL TECHNIQUES

Figure 1 identifies a set of statistical techniques which are relevant for combining collateral data in the context of Landsat modeling and classification. The techniques can be seen as a double level hierarchy, ranging from continuously measured independent variables in the first column to categorically measured independent variables in the last column. In the first row, the dependent variables are continuous in nature, and in the second row the dependent variables are categorical. Continuous variables are measured on interval or ratio scales, whereas categorical variables are measured at nominal or ordinal scales. Categorical variables can be of three types:

- (a) dichotomous (e.g., presence or absence, yes or no)
- (b) unordered polychotomous (e.g., land uses; agriculture, urban, forest, etc.)
- (c) ordered polychotomous (e.g., low runoff, medium, and high runoff)

The statistical techniques in the first row have been thoroughly documented and are commonly used in social science (Blalock, 1972; Graybill, 1961; Morrison, 1967; Winer, 1971). However, there has been relatively little work in the area of applying these approaches to remote sensing. The second row of the figure describes techniques which are generally less well known, but also

**ORIGINAL PAGE IS
OF POOR QUALITY**

include maximum likelihood classification as commonly carried out in remote sensing. The remaining portions of this section discuss the theory and remote sensing application of the techniques identified in Figure 1.

2.1 CONTINUOUS RESPONSE VARIABLE

The statistical techniques of the first row of Figure 1 share one thing in common -- they are all different forms of the basic linear regression model. Consequently, the theory and application of the different types of regression in row one will be very similar. Cell (a) will be examined in detail, and except where explicitly specified, the analysis can be extended to cells (b) and (c).

2.1.1 CELL (a) -- CONTINUOUS EXPLANATORY VARIABLES. The most commonly used method in this cell is regression, which predicts a continuously measured response variable from continuously measured explanatory variables. The model (in vector notation) can be written:

$$\hat{Y} = \beta X + \epsilon$$

where \hat{Y} is the vector of observed dependent variables, β is the vector of unknown parameters, X is the vector of observed independent variables, and ϵ is the vector of errors. (Table I describes the symbols used in this paper.) Typically, the vector β is unknown, and must be estimated from a dataset for which both response and explanatory variables are observed. This estimation is done by the process of ordinary least squares regression (OLS). OLS regression finds the slope and the intercept of a line running through the data which minimizes the variance of $\sum(Y_i - \hat{Y}_i)^2$, or the sum of the squared differences between the observed Y and the predicted \hat{Y} . The vector of predicted \hat{Y} value is calculated by:

$$\hat{Y} = \hat{\beta}X$$

The regression is accomplished by defining the quantity Q equal to $\sum(Y_i - \hat{Y}_i)$ taking the first partial derivatives of Q with respect to the values in the vector β and setting these partials equal to zero. By definition of a parital derivative, a minimum has been found.

The best statistic for measuring the strength of the regression is R^2 . There are several ways of calculating R^2 , but the following is conceptually the simplest. The residual sum of squares (RSS) or the total amount of variance in the model not explained by the regression is calculated by:

$$RSS = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

The RSS , when divided by the total corrected sum of squares in Y (written TSS), gives the proportion of unexplained variance to total variance. R^2 , or the proportion of explained variance, is obtained by subtracting this ratio from one:

$$R^2 = 1 - \frac{RSS}{TSS}$$

One example of how regression could be used within the context of merging Landsat and collateral data is biomass modeling. The regression example used for cell (a) is restricted to continuously measured independent variables. For simplicity, only one spectral channel and one collateral data source are used. Adding other channels and other collateral data sources is a straightforward procedure. A basic linear model could be:

$$\hat{B}_i = \hat{\beta}_0 + \hat{\beta}_1 (MSS_i) + \hat{\beta}_2 (rain_i)$$

where \hat{B}_i is the predicted biomass for pixel i , MSS_i is observed multispectral scanner data (as single band, or multiband ratio or transform) for pixel i , $rain_i$ is observed rainfall on pixel i , and the vector of betas ($\hat{\beta}_0, \dots, \hat{\beta}_2$) are the estimated regression parameters.

Regression models are applied as a two stage process. First, the model must be calibrated (estimate the vector of β 's) by regressing the observed independent variables on the observed

ORIGINAL PAGE IS
OF POOR QUALITY

dependent variable. In this example, biomass data are required from a sufficient number of pixels to give a representative sample of the Landsat image to be modeled. The locations of data points are then "rubber-sheeted" to a geometrically corrected Landsat image, and the linked biomass and MSS data is directly accessed by a statistical software package, such as the Statistical Analysis System (SAS), to calculate the betas and R^2 . Secondly, if R^2 is statistically significant (i.e., the calculated betas explain a significant portion of the total variance in y), the model is extended as a predictor to other pixels where the dependent variable has not been observed. This, in effect, constitutes a model of biomass predicted by surface reflectance and measured rainfall.

2.1.2 CELL (b) -- MIXED EXPLANATORY VARIABLES. This cell includes conventional regression models which are similar to those of cell (a) but also include a mixture of continuous and categorical explanatory variables. Such models are common in social science research, the categorical explanatory variables often being termed "dummy" variables. It can be shown that the more familiar statistical test Analysis of Covariance (ANCOVA) is a straightforward extension of OLS regression with dummy variables (variables that assume values of 1 or 0 depending on the presence or absence of the qualitative variable being measured).

As an example, the model used in cell (a) will be extended to cell (b). In this cell we are able to include data measured categorically -- for example, soil type which can be observed in two states, referred to as class 1 and class 2. The model now becomes:

$$\hat{B}_i = \hat{\beta}_0 + \hat{\beta}_1(MSS_i) + \hat{\beta}_2(rain_i) + \hat{\beta}_3D_{i1} + \hat{\beta}_4D_{i2}$$

where the biomass for pixel i is predicted by the variables used in cell (a) in combination with the dummy variable terms $\hat{\beta}_3D_{i1}$ and $\hat{\beta}_4D_{i2}$. If the observed soil type for pixel i is class 1, then the categorical beta that will be used is $\hat{\beta}_3$, whereas if the soil type is class 2, then $\hat{\beta}_4$ will be used. This is accomplished by defining D_{i1} equal to 1 if the soil type in pixel i is class 1 and 0 otherwise and by defining D_{i2} equal to 1 if the soil type is class 2 and 0 otherwise. It is a relatively simple task to expand this model to include several (polychotomous) soil types or to include other categorical data.

When there is more than one interval scale dependent variable, the model is called MANCOVA, or Multivariate Analysis of Covariance. In this model, all of the independent variables are regressed against each of the dependent variables, with separate R^2 's and F ratios calculated for each dependent variable. An example of this is:

$$\hat{B}_i, \hat{S}_i = \hat{\beta}_0 + \hat{\beta}_1(MSS_i) + \hat{\beta}_2(rain_i) + \hat{\beta}_3D_{i1} + \hat{\beta}_4D_{i2}$$

where \hat{B}_i is predicted biomass for pixel i , \hat{S}_i is predicted soil loss for pixel i , and the other variables remain as in the preceding example. MANCOVA can be seen as a device to test more than one ANCOVA model in the same statistical analysis.

2.1.3 CELL (c) -- CATEGORICAL EXPLANATORY VARIABLES. The extended regression model examined in cell (b) is also applicable here. When the explanatory variables are all measured on the ordinal or nominal scales, the model is usually called Analysis of Variance (ANOVA). All that is necessary to move from cell (b) to (c) is to exclude from the analysis all explanatory variables that are measured on the interval or ratio level. As an example of ANOVA, the biomass model would be written:

$$\hat{B}_i = \hat{\beta}_0 + \hat{\beta}_1D_{i1} + \hat{\beta}_2D_{i2} + \hat{\beta}_3D_{i3} + \hat{\beta}_4D_{i4}$$

Here, $\hat{\beta}_1$ and $\hat{\beta}_2$ refer to two different soil types and $\hat{\beta}_3$ and $\hat{\beta}_4$ refer to rainfall that has been categorized into two levels (high and low). It would be possible to categorize MSS data for use in such a model, but there is considerable information loss. Consequently, the ANOVA or MANOVA (Multivariate Analysis of Variance) model is not likely to be as useful as the ANCOVA or MANCOVA model.

2.2 CATEGORICAL RESPONSE VARIABLE

The first three cells have dealt with Landsat MSS data and continuous dependent collateral variables as predictors in various statistical models of physical phenomena. The last three

**ORIGINAL PAGE IS
OF POOR QUALITY**

cells, in which the dependent variable is categorical, open up new ways of utilizing collateral data within the Landsat structure. With categorical data and with the appropriate statistical techniques (see Figure 1) it is possible to:

- 1) model physical and social phenomena that are best represented in discrete steps -- low, medium, high (soil erosion, fire hazard, biomass, housing quality, municipal services, etc.);
- 2) classify the dependent variable into nominal groupings (land use, vegetation community type, etc.);
- 3) create predicted probabilities that the dependent variable will assume particular categories and use these probabilities to classify an image directly or use them in conjunction with other data (usually MSS) in a Bayesian maximum likelihood classifier.

Unlike the first row of Figure 1, the second row includes five different statistical techniques. For purposes of modeling with Landsat under the constraints of the second row, Maximum Likelihood Classification (MLC) with Prior Probabilities is the most important method. "Maximum likelihood" is a statistical property of an estimator, and, used in its proper way, implies that an estimator has the highest probability of producing the data which were used for calibration. However, its use in remote sensing implies a particular decision rule (MLC) possessing this property, which is discussed below. Our research has shown that the inclusion of collateral data as prior probabilities to MLC offers a simple and effective way of combining collateral and remotely sensed data.

The key to the use of prior probabilities is the logit regression model, which takes collateral data (continuous or categorical) and generates predicted probabilities. These predicted probabilities can be used as a direct classifier (i.e., the pixel is classified into the category that has the highest probability), but the most likely usage of these predicted probabilities is as input to the MLC decision rule, in which they serve as weights. Since the logit regression model is probably the least familiar of the statistical techniques to remote sensing research, and since it is applicable in every cell in the bottom row, it will be explored with the most detail.

Discriminant Analysis is similar in its usage to maximum likelihood with prior probabilities but because of its computational complexity it has not been used often in the remote sensing context. Consequently, it will be only briefly discussed. Chi-Square Analysis, which is the traditional analysis used on contingency tables (all variables are categorical) is by definition not compatible with interval or ratio scale remotely sensed data. It is similar to ANOVA, except that the output is categorical.

2.2.1 CELL (d) -- CONTINUOUS EXPLANATORY VARIABLES. In the past ten years, maximum likelihood classification has found wide application in the field of remote sensing. Based on multivariate normal distribution theory, the MLC algorithm has been in use for applications in the social sciences since the late 1940's. Providing a probabilistic method for recognizing similarities between individual measurements and predefined standards, the algorithm found increasing use in the field of pattern recognition in the following decades (Chow, 1957; Sebestyen, 1962; Nilsson, 1965). In remote sensing, the development of multispectral scanning technology to produce layered multispectral digital images of land areas from aircraft or spacecraft provided the opportunity to use the maximum likelihood classifier in producing thematic classification maps of large areas for such purposes as land use/land cover determination and natural cultivated land inventory (Schell, 1972; Reeves et al., 1975).

Before presenting a practical example, it will be helpful to briefly review the mathematics of the maximum likelihood decision rule. In the multivariate remote sensing application, it is assumed that each observation X (pixel) consists of a set of measurements on p variables (channels). Through some external procedure a set of observations which correspond to a class is identified -- that is, a set of similar objects characterized by a vector of means on measurement variables and a variance-covariance matrix describing the interrelationships among the measurement variables which are characteristic of the class. Although the parametric mean vector and dispersion matrix for the class remain unknown, they are estimated by the sample means and dispersion matrix associated with the object sample.

Multivariate normal statistical theory describes the probability that an observation X will

**ORIGINAL PAGE IS
OF POOR QUALITY**

occur, given that it belongs to a class k , as the following function:

$$\phi_k(X_j) = (2\pi)^{-\frac{1}{2}p} |\Sigma_k|^{-\frac{1}{2}} e^{-(X - \mu_k)' \Sigma_k^{-1} (X - \mu_k)}$$

As applied in a maximum likelihood decision rule, the previous expression allows the calculation of the probability that an observation is a member of each of k classes. The observation (pixel) is then assigned to the class for which the probability density value is greatest. Since the log of the probability is a monotonic increasing function of the probability, the decision can be made by comparing values for each class as calculated from the right hand side of equation.

A simplified remote sensing classification rule using maximum likelihood (Tatsuoka, 1971; Strahler, 1980) with k possible categorical classes and p channels of MSS input datasets is to choose the k (class) which minimizes

$$F_{1,k}(X_j) = \ln |D_k| + (X_j - m_k)' D_k^{-1} (X_j - m_k).$$

This expression is derived from the preceding one by taking the natural logarithm and deleting terms which are constant for all classes.

Interval or ratio level collateral data can be incorporated as extra "logical" channels within this model. One successful forestry application was achieved by the creation of a texture channel which was synthesized from Landsat Band-5 by taking the standard deviation of density values within a 3 by 3 moving window, scaling this value, associating it with the center pixel of the 3 by 3 window, and returning it in image format (as a fifth channel). Values in the texture channel describe the variation in image tone within the immediate area of each pixel. High values are characteristic of edges and boundaries, whereas lower values describe more uniform areas. This technique was shown to significantly increase classification accuracies for a number of species-specific forest cover types in northern California (Strahler, 1978, 1979). Strahler (1978) demonstrated how to input collateral information in the form of elevation data and slope aspect (in combination with a texture channel) as separate "logical" channels, increasing the classification accuracy by 27 percent.

2.2.1.1 Logit Regression. In extending the conventional regression models adopted in cells (a), (b), and (c) to the problems of cell (d), two difficulties are encountered. (For details see Wrigley, 1976, p. 8-9; 1977b; p. 12-13). First, a conventional regression model with a categorical response variable will violate the constant error variance or homoscedasticity assumption. While this problem does not result in biased or inconsistent parameter estimates, it does result in a loss of efficiency and gives rise to serious problems if conventional inferential tests are used. Secondly, a conventional regression model with a categorical response variable may generate predictions which are seriously deficient. It can be shown that the predicted values of the response variable in such a model are best interpreted as predicted probabilities. The problem is that although probabilities are constrained to lie within the range of 0 to 1, the predictions generated from such a model are unbounded and may take values from minus infinity to plus infinity. Thus, the predictions may lie outside the meaningful range of probability and may be inconsistent with the probability interpretation that was just presented. The simplest yet most statistically sound solution to the probability problem (within a regression framework) is the logit transformation, in which the probability P_i is modeled as

$$P_i = \frac{e^{\hat{\beta}X}}{1 + e^{\hat{\beta}X}}$$

and the probability of "not P_i " is:

$$1 - P_i = \frac{1}{1 + e^{\hat{\beta}X}}$$

where $\hat{\beta}X$ is the vector product of betas multiplied by row vector of X 's (observed explanatory variables). Although these two equations are nonlinear models, it is a simple matter to rewrite them as

ORIGINAL PAGE IS
OF POOR QUALITY

$$\frac{P_t}{1 - P_t} = e^{\hat{\beta}X}$$

The logit transformation is achieved by taking the natural logarithm of the preceding formula, which yields

$$\ln \frac{P_t}{1 - P_t} = \hat{\beta}X$$

This transformation has the property of increasing from minus infinity to plus infinity as P_t increases from 0 to 1. Once efficient estimators are calculated for the betas, simple algebra will extract the value of P_t . The method can be generalized to k classes, in which there are $k - 1$ logits of the form

$$\ln(\frac{P_2}{P_1}), \ln(\frac{P_3}{P_1}), \dots, \ln(\frac{P_k}{P_1})$$

Each logit must be modeled separately, producing $k - 1$ sets of betas. As in the binary case described above, algebra will extract the values of the probabilities from the $k - 1$ logits predicted for an observation along with the constraint that all probabilities must sum to one.

Unlike the conventional regression models of the previous cells, all of which can be efficiently estimated by the ordinary least squares (OLS) method, the logistic and linear logit regression models appropriate for the problems of cell (d) require either a weighted least squares (WLS) estimation procedure or a maximum likelihood procedure. The choice between the two methods depends upon whether the calibration data include repeated observations for each combination of values of the explanatory variables (in the case of WLS) or not (in the case of maximum likelihood). Since such replications are unlikely in calibration of the logit model for remotely sensed data, maximum likelihood estimation is preferred. (Note that here the term "maximum likelihood" refers to a parameter estimation method different from multivariate normal MLC.) The maximum likelihood solution to the calibration of the logit model has many attractive features. It can be shown that provided the sample data are not multicollinear, a unique maximum likelihood estimator can be obtained even in relatively small samples. Also, the mathematical properties of the likelihood function allow for efficient computer programs to produce the parameter estimates, and these estimates are consistent and are the best possible estimates in very large samples. The disadvantages of the procedure are that it involves numerical optimization and therefore more costly computation, and that it is a less familiar statistical technique. For a description of the procedure used to calculate such maximum likelihood estimates, see Cox (1970, p. 87), Mantel and Brown (1973, p. 654-5), Wrigley (1975, p. 191-3), Domencich and McFadden (1975, p. 110-12) and Schmidt and Strauss (1975a, p. 404-5).

There are two ways the logit model can be used within the context of remote sensing. The first is to act directly as a nonparametric classifier (i.e., to classify a pixel into the class having the highest predicted probability). The second is to use collateral data in a logit model to predict prior probabilities and input them to a MLC decision rule which accepts prior probabilities. This approach effectively combines a nonparametric logit model for collateral data with a parametric MLC model; it will be discussed in a following section. The following example of the logit model involves the direct classification of land use by MSS and terrain data. The model is

$$\ln \frac{P_{i1}}{1 - P_{i1}} = \hat{\beta}_0 + \hat{\beta}_1(band5_i) + \hat{\beta}_2(slope_i)$$

where P_{i1} is the probability that pixel i is class 1 and the following terms indicate linear combination of spectral and collateral data. If desired, the model can easily be expanded to all four bands and all the continuous collateral variables relevant to the classification.

2.2.1.2 Discriminant Analysis. Discriminant analysis is a multivariate technique used to produce sets of uncorrelated functions which separate observations most efficiently into predesignated groups. A discriminant model, or classification criterion, is developed, the values of

ORIGINAL PAGE IS
OF POOR QUALITY

which define groups for the observations. The individual observation is classified into one of the previously defined groups by a measure of generalized squared distance.

This technique requires some difficult computation. Deriving the classification criteria requires extracting eigenvectors from the nonsymmetric $W^{-1}B$ matrix, where B and W are respectively, the between and within groups sums of squares and crossproducts matrices. The mathematics of this process are beyond the scope of the paper (please refer to Tatsuoka, 1971; Cooley and Lohnes, 1971). The technique is helpful in the social sciences for identifying variables which do the best job of separating classes, but is not usually used to process new data for classification. Further, the technique assumes that all classes possess an identical dispersion matrix, an assumption unlikely to characterize remotely sensed data.

2.2.2 CELL (e) -- MIXED EXPLANATORY VARIABLES. Conceptually no new problems are encountered in moving from cell (d) to cell (e); categorical explanatory variables are included through dummy variables. The logit model has high potential for application in remote sensing. Since it is capable of incorporating both continuous and discrete input data and generating probabilities either directly for classification or as input as a collateral data set of probabilities to MLC (Strahler, 1979). In other words, interval level measured data such as rainfall, elevation, slope, etc. (no matter what its variance-covariance) can be combined with discrete data such as soil type, previously classified land use, census tracts, etc., in a nonparametric, logit framework and the result will be a discrete output such as a land use map, a soil erodibility map (divided into discrete levels) or other special use maps.

The logit model, as in the previous cell, can be easily extended to cell (e). Again, land use could be predicted by

$$\ln \frac{P_{i1}}{1 - P_{i1}} = \hat{\beta}_0 + \hat{\beta}_1(band5_i) + \hat{\beta}_2(slope_i) + \hat{\beta}_3D_{i1} + \hat{\beta}_4D_{i2}$$

where P_{i1} is the probability of land use '1' for pixel i over the probability of all the land uses that are not '1', $band5_i$ is the MSS value for pixel i , $slope_i$ is slope of the pixel, and D_{i1} is a dummy variable with a value of 0 if the previous land use on pixel i was class 1 and 0 otherwise and D_{i2} has the value 1 if the land use on pixel i was class 2 and 0 otherwise.

Preceeding paragraphs have referred to the use of prior probabilities in modifying the outcome of a MLC. Since the prior probabilities can be modeled to reflect both continuous and categorical input data, and the input of MLC is a categorical classification, this technique is appropriate to discussion of cell (e). Prior probabilities are incorporated into the classification through the manipulation of the Law of Conditional Probability. The acutal derivation of the prior probability is beyond the scope of this paper (see Strahler, 1980). The modified decision rule is to choose k which minimizes

$$F_{2,k}(X_j) = \ln|D_k| + (X_j - m_k)'D_k^{-1}(X_j - m_k) - 2\ln P(\theta_k)$$

where the only difference between this formula and the one presented in cell (d) is the probability term, $-2\ln P(\theta_k)$. This form of the decision rule is usually attributed to Tatsuoka and Tiedeman (1954; Tatsuoka, 1971).

It is important to understand how this decision rule behaves with different prior probabilities. If the prior probability $P(\theta_k)$ is very small, then its natural logarithm will be a large negative number; when multiplied by -2, it will become a large positive number and thus $F_{2,k}$ for such a class will never be minimal. Therefore, setting a very small prior probability will effectively remove a class from the output classification. Note that this effect will occur even if the observation vector X_j is coincident with class mean vector m_k . In such a case, the quadratic product distance function $(X_j - m_k)'D_k^{-1}(X_j - m_k)$ goes to zero, but the prior probability term $-2\ln P(\theta_k)$ can still be large. Thus it is entirely possible that the observation will be classified into a different class, one for which the distance function is quite large.

As the prior probability $P(\theta_k)$ becomes large and approaches 1, its logarithm will go to zero and $F_{2,k}$ will approach $F_{1,k}$ for that class. Since this probability and all others must sum to one, however, the prior probabilities of the remaining classes will be small numbers and their

ORIGINAL PAGE IS OF POOR QUALITY

values of $F_{2,k}$ will be greatly augmented. The effect will be to force classification into the class with high probability. Therefore, the more extreme are the values of the prior probabilities, the less important are the actual observations values X_j .

For a numerical example of how prior probabilities can affect the decision of the maximum likelihood classifier, please refer to Strahler, 1980. There are so many potential applications of prior probabilities and the maximum likelihood decision rule that it would be counterproductive to list them all. In general, *all* data that is relevant to a classification model can now be incorporated and this process has been shown to significantly increase classification accuracies (Strahler, 1978; Strahler, 1980).

The versatility of the prior probability techniques comes about when the priors are allowed to vary on a pixel-by-pixel basis. The priors for a pixel may be determined by a logit or other model, or by using a set of class-conditional prior probabilities estimated by sampling. Because the priors are computed separately, it is possible to mix any sort of model estimating prior probabilities with a multivariate normal MLC algorithm which is known to be well suited to most spectral data. Thus, the technique allows easy, flexible merging of collateral data, used to predict the priors, with continuous image data. These points are discussed in more length in Strahler (1978).

2.2.3 CELL (f) -- CATEGORICAL EXPLANATORY VARIABLES. Data that falls into this level has been traditionally analyzed by contingency table analysis with Chi-Square methods. But statistically speaking, it is a simple matter to extend the logit model of cell (e) to cell (f) through the use of dummy variables. For data that only come in nominal or ordinal levels, the logit model offers new and important insights into the data (Wrigley, 1979; Theil, 1970; Grizzle, Starmer and Koch, 1969; Koch et al., 1971, 1972, 1976a, 1977; Landis and Koch, 1977; Lehnen and Koch, 1974a, 1974b). As in earlier discussion, the logit model can serve directly as a nonparametric classifier, using only categorical variables input as dummy variables. In this form, the logit model is equivalent to a log linear model of a contingency table; such models are discussed fully in such texts as Bishop et al. (1975).

The categorical logit model is formulated in the example below:

$$\ln \frac{P_{1i}}{1 - P_{1i}} = \hat{\beta}_0 + \hat{\beta}_1 D_{1i} + \hat{\beta}_2 D_{2i}$$

where there are two output categories -- 1 and not 1 -- which are modeled by categories of soil type as described in cell (e). The classifier simply assigns the output pixel to the class with the higher probability.

3. APPLICATIONS

Two statistical modeling techniques, logit modeling and maximum likelihood classification with prior probabilities, were selected for further investigation in the context of a real application. A linear logit model was devised and fitted to forest species compositional data for northern California, predicting the proportion of timber volume for each of five coniferous species at each pixel based on registered terrain data quantifying elevation and slope aspect. Maximum likelihood classification with prior probabilities was tested in Ventura County, California in a land use application. A previous Landsat classification and an externally devised transition probability matrix were used together with new Landsat image data to produce an updated classification consistent with the observed pattern of change.

Throughout this research we have utilized the Video Image Communication and Retrieval (VICAR) system and the Image Based Information System (IBIS) resident at UCSB. VICAR/IBIS, developed at the Jet Propulsion Laboratory (JPL) at Pasadena, California, USA, is a job control language which permits the sequential linking and execution of a vast array of Fortran and Assembler routines in a batch environment. In addition to extensive usage of existing VICAR/IBIS routines, new VICAR and non-VICAR software were developed and/or modified as required for the purposes of this research.

3.1 LOGIT MODELING OF SPECIES PROPORTIONS

**ORIGINAL PAGE IS
OF POOR QUALITY**

Logit modeling of species proportions used data derived from the Klamath National Forest, located in northern California, USA, (Figure 2). Ranging in relief from 500 to 8,000 feet, the Forest includes 2,600 square miles of rugged terrain in the Siskiyou, Scott Bar, and Salmon Mountains. Little of the area is developed beyond management for timber yield, livestock production, and recreation. A wide variety of distinctive vegetative types is present in the area. Forest vegetation includes such coniferous species as noble, red, white, and douglas firs, ponderosa pine, and incense cedar, as well as several oaks, and typical species of chaparral. Thus, the topographic and vegetational characteristics of the area are well differentiated. Within the Klamath National Forest, a study area including most of the Goosenest Range was selected for logit modeling of species composition from terrain features. This area was chosen because calibration data and Landsat images were readily available for it.

The logit model devised for this forestry application requires preparation of digital terrain data. These data, obtained from the National Cartographic Information Center, in Reston, Virginia, USA, are derived from processing of 1:250,000 contour maps, and include elevations at every point on a grid of approximately 65 m spacing. Although the data are comparable in scale to a Landsat image, the elevation values are quite generalized because they are produced from small scale contour maps by interpolation (Figure 4).

Slope angle and slope aspect channels can be produced using the elevation data of the registered terrain image. Although a number of slope and aspect generating algorithms are known, the simplest is the fitting of a least squares plane through each pixel and its four nearest neighbors and the calculation of the downslope angle and direction of the plane. Slope angle is obtained relative to the numeric range of the elevation channel and image grid spacing, and azimuth is determined with respect to the rectangular image grid. These channels are best generated directly during the preprocessing of the original terrain data. At that time, slope angles and aspects can be calculated from half-word absolute elevations arrayed in a north-south east-west grid.

The slope aspect image (Figure 5) consisted initially of gray tone densities between 0 (black) and 255 (white) which indicated the azimuth of slope orientation, ranging clockwise from 0° to 359°. These values were then transformed according to the function below:

$$\text{newden} = 3.0 + 126.0 * (1.0 + \cos(.024933275 * (\text{oldden} - 26.1)))$$

where *oldden* symbolizes the old (azimuth-keyed) gray tone pixel density value, *newden* symbolizes its transformed value, and the argument of the cosine function is expressed in radians. This function transforms density values according to an orientation proposed by Hartung and Lloyd (1969). Since northeast slopes present the most favorable growing environment, and southwest slopes the least favorable, with northwest and southeast slopes of neutral character, the density tone azimuths were rescaled by a cosine function with 3 representing due northeast and 255 representing due southwest. Neutral slopes, oriented northwest or southwest, thus received density tones near 128. Flat pixels were coded with zeroes. The function also corrects automatically for the 12° skew of the Landsat image.

The logit model fitted is shown below:

$$\ln\left(\frac{P_k}{P_{\sim k}}\right) = \hat{\beta}_{1,k} + \hat{\beta}_{2,k}E + \hat{\beta}_{3,k}A + \hat{\beta}_{4,k}S, \quad k = 1, 5$$

where P_k is the probability that the board-foot of timber volume will be drawn from one of five species k , $P_{\sim k}$ is the probability that the board-foot will not be drawn from species k , E is elevation (compressed to 0-255 range), A is aspect transformed as described above, S is slope angle, and $\hat{\beta}_{1,k}, \dots, \hat{\beta}_{4,k}$ are the estimated regression constants. Note that five equations, one for each species, actually comprise the model. The model was calibrated using 73 measurements of timber volume prepared by the U. S. Forest Service and located within two subregions of the Goosenest range. These samples are probably not representative of the entire area modeled, but serve for the demonstration purposes of this research. Each sample was located on 1:15,840 scale color air photos and transferred to Band 5 of the Landsat images to obtain the line and sample coordinates of the sample point. The coordinates were then used to extract elevation and aspect values for the sample from the registered elevation and aspect images. The coefficients for the model were fitted by a nonlinear optimization algorithm employing the Newton-Raphson method to select

ORIGINAL PAGE IS
OF POOR QUALITY

coefficients with maximum likelihood.

Given the constants produced by the procedure discussed above, the probability images were created using the new VICAR program "PROBMAPS." PROBMAPS, written specifically for this application, calculates the probability of species k for each pixel using the following expression:

$$P_k = \frac{Q_k}{\sum_{j=1}^5 Q_j} \text{ where } Q_k = \exp(\hat{\beta}_{1,k} + \hat{\beta}_{2,k}E + \hat{\beta}_{3,k}A + \hat{\beta}_{4,k}S)$$

PROBMAPS then scales each probability so that the range 0-255 represents 0. to 1. PROBMAPS output images for this example are shown in Figures 6-10. Brightness values in Figures 6-9 represent probabilities of occurrence for douglas fir, ponderosa pine, white fir, and red fir, respectively, with probabilities scaled to range from black (0.) to white (1.0). Figure 10, incense cedar, has been contrast stretched for display purposes, and presents a probability range of 0. to .3 from black to white. The probability images represent maximum likelihood estimates of species proportions; they appear reasonable in light of the known ecological preferences of the species, but their accuracies remain to be determined.

The probability images produced by PROBMAPS can be thought of as predictions of the proportion of the timber volume expected for each species at each pixel. This view implies a continuous mixture of species, constantly varying in response to elevation and slope aspect. An alternative view is that forest stands are monospecific, and that each pixel is dominated by a particular species or forest cover type. In such a case, the modeled values are probabilities that the pixel will be dominated by a particular species or stand type, and it is therefore appropriate to produce a single output image indicating the type with highest probability for each pixel. In this way, the logit model can serve as a nonparametric classifier. The probabilities can also be viewed as prior probabilities, and input to MLC of an image using spectral data. This procedure amounts to mixing a nonparametric model for collateral data (terrain channels) with a parametric model for spectral data (Landsat channels).

3.2 LAND USE CLASSIFICATION USING TRANSITION PROBABILITIES

An additional objective of our research was to apply the method of maximum likelihood classification with prior probabilities to a land use/land cover classification (see sections 2.2.1 and 2.2.2). In this example, land use/land cover maps for two $7\frac{1}{2}$ -minute quadrangles in Ventura County, California, U.S.A., were obtained from photointerpretation of high-level U-2 aircraft imagery for the years 1973 and 1976. These maps, with inherent accuracies considerably higher than those of Landsat classifications, were used as "ground truth" to construct a matrix of transition probabilities, showing the probability of change of classification for each land use/land cover type to each other type in the three year interval. With a 1976 Landsat image as input data, we planned to carry out MLC of each pixel using the 1973 U-2 derived cover class as a collateral data channel indexing the transition probabilities appropriate to the 1973 cover class. The resulting classification was to be compared with the 1976 U-2 derived map to evaluate the accuracy of the technique. At present, we have used image overlay techniques to create the transition probability matrix, but the classifications using the transition probability matrix have not been carried out.

3.2.1 PROCEDURE. Using the Jet Propulsion Lab's (JPL) Image Based Information System (IBIS) and a coordinate digitizer, it is possible to merge image data in digital form with other types of geographic data. The IBIS is essentially a fine-mesh grid information system which is compatible with the handling and storing of digital image data. By allowing a user to overlay thematic map and digital Landsat data (or pertinent Landsat-derived data) with IBIS, it is possible to derive the values that comprise the transition probability matrix, as well as determine the accuracy of the thematic Landsat classification data.

Prior to IBIS processing, the two "ground truth" land use/land cover maps for the two-quadrangle study area were prepared through photointerpretation of NASA high-altitude color infrared imagery with additional ground checks. A coordinate digitizer board was used to convert the maps to a series of digital coordinates. Polygons of thematic land cover categories were captured by digitizing overlapping line segments that comprise such polygons. Polygon centroids were also

ORIGINAL PAGE IS
OF POOR QUALITY

digitized and assigned an appropriate land use/land cover label for later use in converting the polygonal data to raster (image base) form.

The digitized line segment data were processed using IBIS as follows. A modified version of the IBIS program POLYGEN converted the coordinate digitizer segment data into polygons in the form of an IBIS graphics file. Following this reformatting, the program POLYREG rigidly rotated the polygons to conform with the Landsat data for the test site. POLYSCRB was used to convert the polygon data into raster form (fine-mesh grid). The result was an image of rasterized polygon borders representing the edges to thematic land cover units.

The next phase in the IBIS processing involved the assignment of labels to the rasterized polygon areas. An intermediate categorized image was automatically produced by the program PAINT, which assigns an arbitrary but unique brightness member to all the pixels within each rasterized polygon. This output image was combined with the polygon centroids that had also been converted to an IBIS graphics file format (with V2POLY) and rigidly transformed to overlay with the polygon borders (POLYREG). The program CTRMATCH was used to establish the correspondence between centroid labels and polygon brightness numbers, and the program STRETCH was used to reassign identical brightness values to polygons with similar labels, yielding a raster format image corresponding directly with the original land use/land cover map.

With both land use/land cover images in digital image format, the next step is to overlay them using POLYOVLY to analyze the change occurring between the two dates in order to estimate the transition probability matrix. The output of POLYOVLY is a table counting the number of pixels in each combination of classes across the two images. If $S_{i,j}$ denotes the count of pixels in class i of the first image, and class j of the second, then the transition probability $T_{i,j}$ for class i to j is

$$T_{i,j} = \frac{S_{i,j}}{\sum_{i=1}^k S_{i,j}}$$

used to assess the accuracy of MLC of Landsat data, again using the POLYOVLY program. Accuracies discussed below were obtained in this fashion.

With a digitized 1973 land use/land cover classification map produced from air photo interpretation now in hand as collateral data channel registered to the 1977 Landsat image, and with a transition probability matrix to provide sets of prior probabilities contingent as a collateral data channel, it should be possible to carry out maximum likelihood classification using the transition probabilities as prior probabilities indexed by the 1973 classification. Future work includes performing the classification, and overlaying it on the 1976 digitized map for accuracy analysis. Initial indications are that accuracies will increase with the use of the 1973 digitized map data, demonstrating the successful use of Landsat to update existing manually produced land use/land cover maps.

4. CONCLUSION

Of the large number of statistical techniques which can be used to develop models combining remotely sensed images with collateral data in a common predictive framework, two techniques are of special interest for remote sensing: logit modeling and maximum likelihood classification with prior probabilities. These methods allow the construction of nonparametric classification models utilizing both image and collateral data channels as well as the mixing of parametric and nonparametric classification models for image and collateral data respectively. Both techniques have been successfully demonstrated in applications using Landsat imagery; each has the potential to greatly increase classification accuracy through the use of collateral data, and each should find wide application in future research and development in remote sensing.

5. REFERENCES

- Bishop, Y.M., Fienberg, S.E., and Holland, P.W. (1975), Discrete Multivariate Analysis: Theory and Practice, MIT Press, Cambridge, Massachusetts.
- Blalock, H.M. (1972), Social Statistics, McGraw-Hill Inc., New York.
- Chow, C.K. (1957), An optimum character recognition system using decision functions, IREE Trans. Electron. Computers 6: pp. 247-254.

ORIGINAL PAGE IS
OF POOR QUALITY

- Cox, D. R. The analysis of binary data. 1970. London, Methuen.
- Domencich, T.A. and McFadden, D. Urban travel demand: a behavioral analysis, 1975, Amsterdam, North-Holland.
- Graybill, F.A. (1961), An Introduction Linear Statistical Models, Vol.1, McGraw-Hill Book Co., New York.
- Grizzio, J.E., Starmer, C.F. and Koch, C.G. Analysis of categorical data by linear models. Biometrics 25, 1969, pp. 489-504.
- Hartung, R.F., Lloyd, W.J. (1969), Influence of Aspect on Forests of the Clarksville Soils in Dent County, Missouri, Journal of Forestry, 67, pp. 178-182.
- Koch, G.G., Imrey, P.B. and Reinfurt, D.W. Linear model analysis of categorical data with incomplete response vectors. Biometrics 28, 1972, pp. 663-92.
- Koch, G.G., Freeman, J. L. and Lehnken, R.G. A general methodology for the analysis of ranked policy preference data. International Statistical Review 44, 1976a, pp. 1-28.
- Koch, G.G., Landis, J.R., Freeman, J.L., Freeman, D.H. and Lehnken, R.G. A general methodology for the analysis of experiments with repeated measurement of categorical data. Biometrics 33, 1977, pp. 133-58.
- Landis, J.R. and Koch, G.G. The measurement of observer agreement for categorical data. Biometrics 33, 1977, pp. 159-74.
- Lehnken, R.G. and Koch, G.G. A general linear approach to the analysis of nonmetric data: applications for political science. American Journal of Political Science 18, 1974c, pp. 283-313.
- Lehnken, R.G. and Koch, G.G. The analysis of categorical data from repeated measurement research designs. Political Methodology 1, 1974b, pp. 103-23.
- Mantel, N. and Brown, C. A logistic re-analysis of Ashford and Snowden's data on respiratory symptoms in British coal miners. Biometrics 29, 1973, pp. 649-65.
- Morrison, (1967), Multivariate Statistical Methods, McGraw-Hill Book Co., New York.
- Nilsson, N.S. (1965), Learning Machines - Foundations of Trainable Pattern - Classifying Systems, McGraw-Hill Book Co., New York.
- Reeves, R.G., Anson, A., and Landen, D. (1975), Manual of Remote Sensing, Amer. Soc. of Photogrammetry, Falls Church, VA, 2 vols., 2144 pp.
- Schell, J.A. (1973), in Remote Sensing of Earth Resources, Volume I (F. Shahrokh, Ed.), University of Tennessee Space Institute, Tullahoma, TN, pp. 374-394.
- Schmidt, P. and Strauss, R.P. The prediction of occupation using multiple logit models. International Economic Review 16, 1975b, pp. 471-86.
- Sebestyen, G. (1962), Decision-Making Processes in Pattern Recognition, MacMillan, New York.
- Strahler, Alan H., T.L. Logan, and N.A. Bryant (1978). Improving forest cover classification accuracy from Landsat by incorporating topographic information: Proceedings of the Twelfth International Symposium on Remote Sensing of the Environment, pp. 927-942.
- Strahler, Alan H., and T.L. Logan, and C.E. Woodcock (1979). Forest classification and inventory system using Landsat, digital terrain, and ground sample data: Proceedings of the Thirteenth International Symposium on Remote Sensing of the Environment, pp. 1541-1557.

ORIGINAL PAGE IS
OF POOR QUALITY

- Strahler, Alan H. (1980). The use of prior probabilities in maximum likelihood classification of remotely sensed data: submitted for publication.
- Tatsuoka, M.M., Multivariate analysis: Techniques for Education and Psychological Research. 1971, John Wiley & Sons, New York.
- Tatsuoka, M.M. and Tiedeman, D.V. Discriminant Analysis Review of Education Research, 25, 1954, pp. 402-420.
- Winer, B.J. (1971), Statistical Principles in Experimental Design, 2nd ed., McGraw-Hill Book Co., New York.
- Wrigley, N. (1975), Analyzing multiple alternative dependent variables. Geographical Analysis 7, pp. 187-95.
- Wrigley, N. (1976), An introduction to the use of logit models in geography. Concepts and techniques in modern geography, 10. Norwich: Geo Abstracts Ltd.
- Wrigley, N., (1976b), Probability surface mapping: a new approach to trend surface mapping. Transactions of the Institute of British Geographers, New Series 2, pp. 129-40.
- Wrigley, N. (1979), Development in the statistical analysis of categorical data -- a review. Progress in Human Geography, 3, pp. 315-355.

Table I. Notation

TERM	DEFINITION
\hat{Y}	Vector of estimated dependent variables; where $\hat{\cdot}$ signifies a vector and $\hat{\cdot}$ signifies an estimator.
$\hat{\beta}, \hat{\epsilon}$	Regression notation; vector of estimated betas and vector of observed error terms.
p	Number of measurement variables used to characterize each object or observation.
X	A p -dimensional random vector.
X_i	Vector of measurements on p variables associated with the i th object or observation: $i=1, 2, \dots, N$.
θ_k	Member of the k th set of classes Θ ; $k=1, 2, \dots, K$.
$P(\theta_k)$	Probability that an observation will be a member of class θ_k ; prior probability for class θ_k .
$\pi_k(X_i)$	Probability density value associated with observation vector X as evaluated for class k .
μ_k	Parametric mean vector associated with the k th class.
m_k	Mean vector associated with a sample of observations belonging to the k th class; taken as an estimator of μ_k .
S_k	Parametric p by p dispersion (variance-covariance) matrix associated with the k th class.
D_k	p by p dispersion matrix associated with a sample of observations belonging to the k th class; taken as an estimator of S_k .

**ORIGINAL PAGE IS
OF POOR QUALITY**

INPUT CHANNELS [Independent Variable(s)]

OUTPUT CHANNEL
[Dependent Variable]

	Continuous	Mixed	Categorical
Continuous	Regression Models - linear - curvilinear	Analysis of Covariance Multivariate Analysis of Covariance	Analysis of Variance Multivariate Analysis of Variance
Categorical	Maximum Likelihood Classification Logit Modeling Discriminant Analysis	Maximum Likelihood Classification with Prior Probabilities Logit Modeling	Contingency Table Analysis Logit Modeling

Figure 1. Techniques for Combining Continuous and Categorical Data
(modified from Wrigley, 1979).

ORIGINAL PAGE IS
OF POOR QUALITY

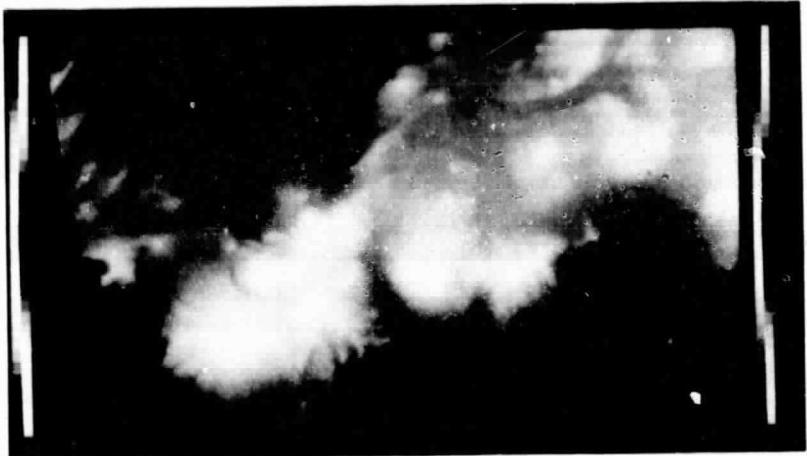


Figure 4. Registered elevation image (derived from N.C.I.G. 1:250,000 digital terrain data).

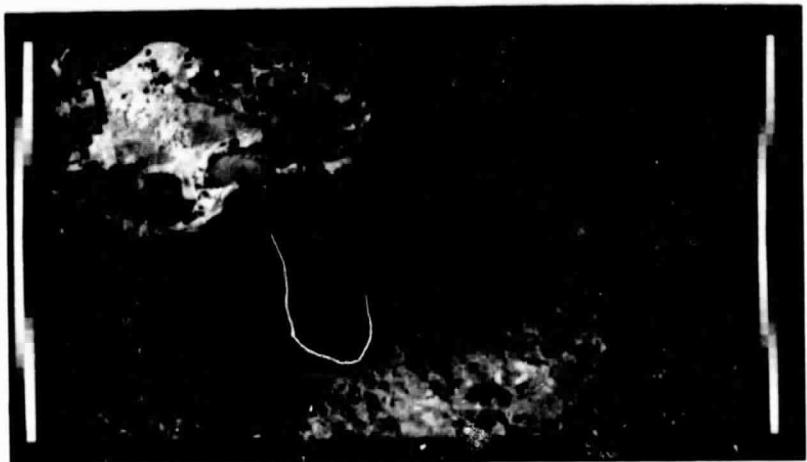


Figure 3. Landsat MSS band 5, Goose-nest test area, Klamath National Forest, California, U.S.A. Scale: pixel = 80 m x 80 m. Note: North is 9° counterclockwise from up.

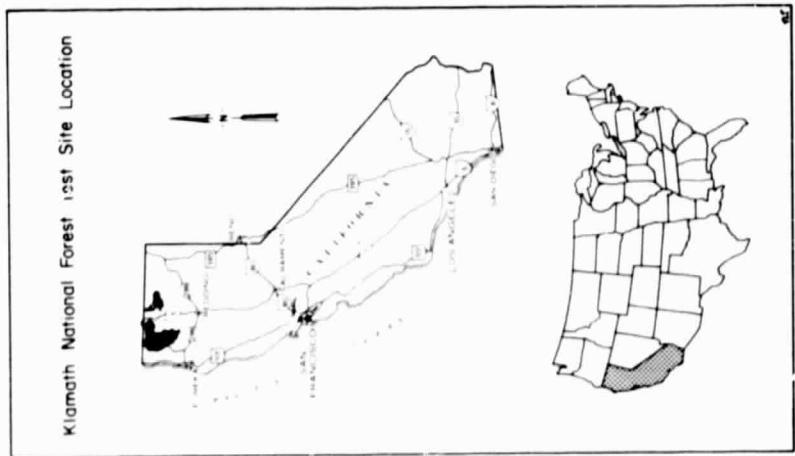


Figure 2. Klamath National Forest location map.

ORIGINAL PAGE IS
OF POOR QUALITY



Figure 7. Registered ponderosa pine probability image.



Figure 6. Registered douglas fir probability image. (Area shown is restricted to National Forest lands).

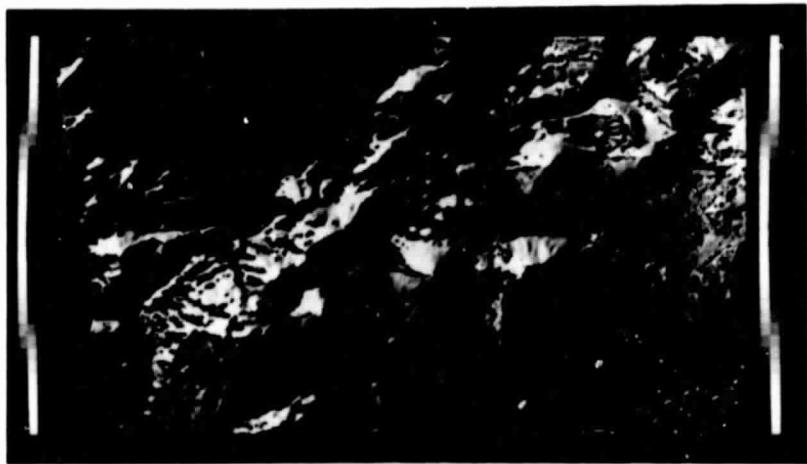


Figure 5. Registered slope orientation image (cosine transformation).

ORIGINAL PAGE IS
OF POOR QUALITY



Figure 10. Registered incense cedar probability image (stretched for contrast enhancement).



Figure 9. Registered red fir probability image.



Figure 8. Registered white fir probability image.

A P P E N D I X C

ORIGINAL PAGE IS
OF POOR QUALITY.

A LOGIT CLASSIFIER FOR MULTI-IMAGE DATA.

Alan H. Strahler and Paul F. Maynard

Department of Geography
University of California
Santa Barbara, California 93106

A new classifier for multi-image databases uses maximum likelihood estimation of parameters fitting a logit model to training data. A logit is the natural logarithm of a probability ratio: e.g., $\ln(P_a/P_b)$. As an example, a linear logit classification model for a simple two-class case based on four Landsat channels is:

$$\ln(\frac{P_a}{P_b}) = \beta_0 + \beta_1(LS4) + \beta_2(LS5) + \beta_3(LS6) + \beta_4(LS7)$$

where P_a and P_b are probabilities that the pixel belongs to class a and b respectively, $\beta_0 \dots \beta_4$ are calibration constants, and LS4...LS7 and the four Landsat channels for bands 4 through 7.

Compared with usual Bayesian maximum likelihood classification, the logit classifier has certain distinct advantages. It is nonparametric, in that multivariate normality is not assumed. The model may be specified in linear or curvilinear forms as appropriate. Further, the model can incorporate categorical information in the form of dummy variables, and can therefore be used to merge continuously measured image data with categorical collateral data in a single classification step.

Introduction

The Bayes maximum likelihood classifier (MLC) is the most commonly used decision rule for discrete classification with Landsat derived data. Such a classifier uses the Bayes decision rule to assign pixels to the class with highest probability, given the observed vector of spectral measurements and the prior distribution of classes. To find this probability, the Bayes MLC requires an estimate of the conditional probability of occurrence of the observed vector of MSS data given that it is associated with a specified class. This estimate has traditionally been obtained by assuming that the observed measurement vectors were Gaussian, or normal; therefore, the best estimate is a measure of the probability density value of the multivariate normal distribution for the class evaluated at the observation.

With four channels of Landsat spectral data, the Bayes maximum likelihood classifier has achieved good classification accuracies in many cases. Classification accuracies have been further boosted by the inclusion of collateral data as additional logical channels or as indexes to sets of prior probabilities in the case of categorical collateral

variables.¹ Further increases in classification accuracy will undoubtedly result from more optimal spectral channels and improved techniques of incorporating collateral data.

However, it is very likely that the spectral reflectances from some classes in many applications are not normally distributed. In this situation, even with the ideal spectral "windows" and precise and relevant collateral data, the Bayes MLC will reach an asymptotic accuracy limit that will be less than optimal. In fact, for classes that deviate from normality, classification accuracies could be significantly less than optimal.

Furthermore, each of the two previously mentioned techniques of increasing accuracy by input of collateral data to the Bayes classifier has critical limitations. The Bayes MLC can only accept measurement variables which are continuously distributed, and this requires continuous measurement (or at least discrete measurement with a large number of discrete steps). Consequently, the collateral channel in the direct input approach can utilize only data which are measured on the interval or ratio scale. This requirement eliminates many potentially useful databases, such as soils maps, geologic maps, political boundaries, census tracts, etc. And when collateral information is incorporated through the mechanism of prior probabilities, the calculation of the prior probabilities usually requires a sophisticated sampling design.

One solution to these problems is to use a statistical technique that predicts probabilities of categorical membership with no distribution assumptions. One such technique, called the logit regression model, has been widely used in the social sciences for the last twenty years. The logit regression model generates predicted probabilities that all sum to one for a specified suite of classes, and the classification can be awarded to the category with the highest predicted probability. Further, predictor variables may be continuous or categorical; the model may be specified in linear or curvilinear forms; and the assumption of multivariate normality is not required.

This paper has the following components:

1. a review of the mathematics of the Bayes MLC;
2. an examination of the sources of classification error due to non-normal distributions;

Reprinted from PROCEEDINGS OF THE WORKSHOP ON
PICTURE DATA DESCRIPTION AND MANAGEMENT, August 1980

ORIGINAL PAGE IS
OF POOR QUALITY

3. an examination of the problems encountered with utilizing collateral data;
4. development and exposition of the logit regression model, with an emphasis on its ability to act as a nonparametric classifier;
5. a description of planned research which will compare the classification accuracies of the logit regression model and the Bayes MLC for a land use/land cover classification example; and
6. presentation of an example of the use of linear logit model in a remote sensing application.

The Bayes Maximum Likelihood Classifier

Background

In the past ten years, maximum likelihood classification has found wide application in the field of remote sensing. Based on multivariate normal distribution theory, the maximum likelihood classification algorithm has been in use for applications in the social sciences since the late 1940's. Providing a probabilistic method for recognizing similarities between individual measurements and predefined standards, the algorithm found increasing use in the field of pattern recognition in the following decades.^{2,3,4} In remote sensing development of multi-spectral digital images of land areas from aircraft or spacecraft provided the opportunity to use the maximum likelihood criterion in producing thematic classification maps of large areas for such purposes as land use/land cover determination and natural cultivated land inventory.^{5,6}

Derivation of the Bayes MLC

In order to understand the difference between a classification awarded on the basis of logit-generated predicted probabilities and posterior probabilities derived from the Bayes MLC, it will be helpful to briefly review the mathematics of the Bayes maximum likelihood decision rule. In the multivariate remote sensing application, it is assumed that each observation X (pixel) consists of a set of measurements on p variables (channels). Through the selection of training sites, a set of observations which correspond to a class is identified -- that is, a set of similar objects characterized by a vector of means on measurement variables and a variance-covariance matrix describing the interrelationships among the measurement variables which are characteristic of the class. Although the parametric mean vector and dispersion matrix for the class remain unknown, they are estimated by the sample means and dispersion matrix associated with the object sample.

Multivariate normal statistical theory describes the conditional probability that an observation X will occur, given that it belongs to a class k , as the following function:

$$P(X|\omega_k) = (2\pi)^{-\frac{1}{2}} P |\Sigma_k|^{-\frac{1}{2}} e^{-\frac{1}{2}(X-\mu_k)' \Sigma_k^{-1} (X-\mu_k)} \quad (1)$$

(Please refer to Table 1 for a description of the mathematical symbols). As applied in a maximum likelihood decision rule, expression (1) allows the calculation of the conditional probability that an

observation is a member of each of k classes. However, the actual probability desired is the posterior probability $P(\omega_k|X)$; it can be shown⁴ that:

$$P(\omega_k|X) = \frac{\phi_k(x) \cdot P(\omega_k)}{\sum_{k=1}^K \phi_k(x) \cdot P(\omega_k)} \quad (2)$$

This expression leads to the decision rule:

Choose k which minimizes

$$\ln |D_k| + (x - m_k)' D^{-1} (x - m_k) - 2 \ln P(\omega_k). \quad (3)$$

In usual practice the prior probabilities $P(\omega_k)$ are assumed equal, or $P(\omega_k) = 1/R$ where R is the number of classes. In this case, the last term in expression (3) is constant over all R classes, and need not be considered in the decision rule. This equal priors decision rule is used in the currently distributed versions of LARSYS and VICAR, two image processing systems authored respectively by the Laboratory for Applications of Remote Sensing at Purdue University and the Jet Propulsion Laboratory of California Institute of Technology at Pasadena.

Classification Accuracy and the Assumption of Normality

In a typical supervised classification, training sites are selected by the analyst to typify each class. Histograms of spectral values for classes are inspected for multivariate normality, and when a class actually consists of several distinctive signatures, training sites are reaggregated into subclasses, each of which is approximately multivariate normal. In this way, a set of multivariate normal dispersion patterns are defined for the desired classes. It is important to realize that such dispersions, because they are selected to be as "pure" as possible, are probably underdispersed with respect to the true information class. This effect produces a difficulty in the classification of mixed pixels. Since the MLC model does not provide for mixed pixels, the implicit assumption is that mixed pixels are to be classified according to the most probable signature match; the components of the signature which are reflected from the less important classes contained within the pixel are thus regarded as random noise. The mixed pixel, then, is typically classified by comparing probability densities within the tails of overlapping multivariate normal distributions. The accurate classification of mixed pixels under MLC thus requires a good fit of the tails to multivariate normality; however, it is obvious that the selection of training sites for purity will of necessity produce a poor fit in the tails. And, mixed pixels will constitute a large portion of the scene -- up to forty percent in some agricultural applications (F. Hall, personal communications).

This reasoning naturally leads to the consideration of nonparametric classifiers. Brooner et al.⁷ compared the Bayes MLC to a classifier which used $P(X|\omega_k)$ as directly estimated by a sampling procedure, and reported a four percent increase in classification accuracy over MLC. However, direct estimates of $P(X|\omega_k)$ require more data as well as

ORIGINAL PAGE IS
OF POOR QUALITY

dealing with p -dimensional table storage problem. The logit model, discussed in the following pages, provides an alternative which should require fewer data to calibrate and can, by virtue of its curvilinear modeling, approximate the real distribution without assuming multivariate normality.

Problems with Incorporating Prior Probabilities

As shown earlier, the Bayes MLC can easily be modified to take into account prior probabilities which describe how likely a class is to occur in the population of observations as a whole. The prior probability itself is simply an estimate of the proportion of pixels which will fall into a particular class. These prior probabilities are sometimes termed weights, since the modified classification rule will tend to weight more heavily those classes with higher prior probabilities. Strahler¹ showed via simplified numerical examples how these different weights can affect the decision of the Bayes MLC. As the prior probability $P(w_i)$ in expression (3) becomes large and approaches 1, its logarithm will go to zero and the classification decision will effectively be made with expression (1).

However, since this possibility and all others must sum to one, the prior probabilities of the remaining classes will be small numbers, thus increasing the value of the expression. Since the classifications is awarded to the class with the smallest value, the effect will be to force classification into the class with high probability. Therefore, the more extreme are the values of the prior probabilities, the less important are the actual observation values x .

Strahler¹ has demonstrated how prior probabilities can be used as a mechanism to incorporate collateral data in categorical form into the Bayes MLC. His mechanism uses a set of prior probabilities estimated for each collateral category by an external sampling procedure, with the classification algorithm accessing the appropriate set of probabilities contingent upon the collateral category of the pixel. In this fashion, categorical collateral information is merged with multivariate normal information concerning the spectral signatures. Although this approach was proven effective for a forestry application, estimations of the sets of prior probabilities may require considerable data collection, depending on the number of classes and collateral categories. In addition, multivariate normality of signatures is assumed, and the comments of preceding paragraphs apply. In contrast, use of the logit classification model allows categorical and continuous information to be mixed freely through the mechanism of dummy variables. And, again, the model can be fitted in a linear or curvilinear fashion as desired, avoiding the assumption of multivariate normality. Thus, the logit model offers a more natural, straightforward way of incorporating categorical variables into the classification procedure.

The Logit Regression Model

Linear Modeling of Probabilities

The most commonly used predictive multivariate statistical technique is probably ordinary least squares (OLS) regression. The prediction, or estimated value of the dependent variable, is a function of the vector of estimated betas ($\hat{\beta}$) in combination with the vector of observed independent variables (x). The betas are estimated in such a way that the variance about the least squares regression line is minimized.

When used to model probabilities, OLS regression has one major drawback: although probabilities are constrained to lie within the range of 0 to 1, the predictions generated from such a model are unbounded and may take values from minus infinity to plus infinity. Thus, the predictions may lie outside the meaningful range of probability. Further, the probability of each class must be modeled separately, and there is no constraint to ensure that all probabilities must sum to one.

One solution to the bounding problem is to specify that

$$0 \leq P_i \leq 1$$

(where P_i is the probability of observing a specified class or category of the dependent variable). In the case of ordinary least squares regression model,

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2}.$$

The simplest way to satisfy this condition is to impose the following arbitrary definition of P_i :

1. P_i is equal to 0 if \hat{Y}_i is less than 0;
2. P_i is equal to \hat{Y}_i if \hat{Y}_i is equal to or between 0 and 1;
3. P_i is equal to 1 if \hat{Y}_i is greater than 1;

and use straightforward ordinary least squares estimation of the regression parameters. This solution is often referred to as the linear probability model. Unfortunately, although it appears to be a simple solution to the predicted probabilities problem, the model has a number of serious limitations which are discussed by Domencich and McFadden.³ Again, the probabilities are not constrained to sum to one.

The Logit Model

The simplest, yet most statistically sound, solution to the probability problem (within a regression framework) is the logit transformation. In this transformation, the ratio between the probability that an observation or pixel i belongs to a class P_1 and the probability that it does not belong to P_1 is expressed as a logistic function:

$$\frac{P_{i,1}}{1-P_{i,1}} = e^{\beta_1 x} \quad (4)$$

ORIGINAL PAGE IS OF POOR QUALITY

where β is a vector of parameters and X is a vector of observations on independent variables. Taking the natural logarithm of this expression,

$$\ln \left(\frac{P_{i,1}}{1-P_{i,1}} \right) = \beta_1 X . \quad (5)$$

The left-hand quantity is referred to as a logit. Note that when βX is zero, the ratio will be 1, indicating equal probability. As βX varies positively or negatively, the ratio will shift accordingly.

The ratio, under the constraint that the numerator and denominator must sum to one, determines the two probabilities uniquely. Expression (4) can be solved explicitly for $P_{i,1}$:

$$P_{i,1} = \frac{e^{\beta_1 X}}{1+e^{\beta_1 X}} . \quad (6)$$

And, if $P_{i,2}$ is defined as $1-P_{i,1}$, it is easy to show that

$$P_{i,2} = 1-P_{i,1} = \frac{1}{1+e^{\beta_1 X}} . \quad (7)$$

Although expressions (4) and (5) show the product βX as a linear function, the β vector may contain powers and cross products in the case of a curvilinear model. An example is (elemental notation):

$$\ln \left(\frac{P_{i,1}}{1-P_{i,1}} \right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{1,1} X_1^2 + \beta_{2,2} X_2^2 + \beta_{1,2} X_1 X_2 \quad (8)$$

for the bivariate case.

Unlike the conventional regression models, the logistic and linear logit regression models require either a weighted least squares (WLS) procedure or a maximum likelihood procedure to estimate the calibration parameters (betas). The choice between the two methods depends upon whether or not the sample under investigation includes repeated observations for each combination of values of the explanatory variables. If so, WLS is appropriate; however, remote sensing applications rarely have repeated observations. Consequently, the method of maximum likelihood is the preferred method. A number of authors present the details of this method, which is discussed briefly in a following section.^{8,9,10,11,12}

Maximum likelihood estimation of logit model parameters has many other attractive features. Provided that the sample data are not multicollinear, a unique maximum likelihood estimator can be obtained even in relatively small samples. Also, the mathematical properties of the likelihood function allow for efficient computer programs to produce the parameter estimates. These estimates are consistent and are the best possible estimates in very large samples. The disadvantages of the procedure are that it involves numerical optimization and therefore more computation, and that it

requires more calibration data than MLC because of the larger number of parameters which need to be estimated.

Logit Example

The following is an example that uses continuous and categorical explanatory variables to estimate

$$\ln \left(\frac{P_{i,1}}{1-P_{i,1}} \right) = \beta_0 + \beta_1 X_{i,1} + D_1 \beta_2 , \quad (9)$$

where $P_{i,1}/(1-P_{i,1})$ is the ratio of the probability that pixel i is not of class 1, $X_{i,1}$ refers to a continuously measured variable on pixel i (MSS or continuous collateral data), and D_1 is a dummy variable that is equal to one if a categorical variable is true at pixel i and zero if it is not. Given the logit ratio, simple algebra will extract the value of $P_{i,1}$. It is straightforward to add more continuous and categorical explanatory variables.

Estimating the Regression Parameters

In order to calculate the logit ratio in the preceding formula, it is necessary to obtain estimates of the regression parameters. The first step is to specify the model in terms of a likelihood function. If the training observations are thought of as independent trials, then the likelihood of the outcome of these trials (for the two-class case described above) is:

$$L = \prod_{i=1}^{n_1} P_{i,1} \prod_{i=n_1+1}^n (1-P_{i,1}) \quad (10)$$

where observations $i=1$ to n_1 are those in which the observed dependent variable was a member of class 1 and $i=n_1+1$ to n are the observations in which the dependent variable was not a member of class 1. Substituting from the definitions of $P_{i,1}$ and $1-P_{i,1}$ in expressions (6) and (7), the result is

$$L = \prod_{i=1}^{n_1} \frac{e^{\beta X_i}}{1+e^{\beta X_i}} \prod_{i=n_1+1}^n \frac{1}{1+e^{\beta X_i}} . \quad (11)$$

As specified, the likelihood depends upon a set of unknown parameters, the betas. These parameters are estimated by choosing those values which maximize the preceding likelihood formula for the given set of training data. Rather than maximize the likelihood itself, it is computationally simpler to maximize the logarithm of the likelihood, or

$$\ln L = \sum_{i=1}^{n_1} \beta X_i - \sum_{i=1}^n \ln(1+e^{\beta X_i}) . \quad (12)$$

To maximize this expression it is not possible to set the partial derivatives of $\ln(L)$ with respect to the betas to zero, and solve simultaneously for the betas in a direct fashion. Instead, the solution must be obtained by iteratively recalculating $\ln L$ for successive estimates of betas until the partial derivatives converge upon zero.

ORIGINAL PAGE IS
OF POOR QUALITY

There are several mathematical techniques for iteratively converging the first partial derivatives to zero. One well-known technique is the Newton-Raphson Method, which calculates deltas (amount of change) for the betas by forming the matrix of second partial derivatives, inverting it, and postmultiplying it by the vector of the first partial derivatives. That is,

$$\vec{d} = \Sigma^{-1} \vec{G},$$

where \vec{d} is the vector of calculated deltas, Σ^{-1} is the inverted second partial derivative matrix, and \vec{G} is the vector of first partial derivatives. The deltas are subtracted from the betas, and the second partial derivative matrix and first partial derivatives are recalculated with the new betas. The process continues until the vector of first partial derivatives has converged upon zero, at which point the most likely vector of betas has been identified.

Given these maximum likelihood estimates of the betas, the last step in a logit model classification sequence is to use expressions (6) and (7) to calculate the vector of probabilities for each pixel and award the classification to the category with the largest predicted probability.

Polychotomous Logit Regression

The preceding example, although conceptually straightforward, is not applicable when there are more than two categories to be predicted. The dichotomous logit can be easily extended to the polychotomous logit. Now, instead of two categories of interest, there are R possible categories. The model now becomes:

$$P_{i,r} = \frac{e^{\beta_r X_i}}{\sum_{s=1}^R e^{\beta_s X_i}}, \quad (13)$$

where $P_{i,r}$ is the probability that pixel i belongs to the r th category.

Because of the constraint that the probabilities must sum to one, only $R-1$ sets of betas and probability ratios need to be determined. Introducing this constraint on the R th class, it is easy to show algebraically that

$$P_{i,R} = 1 - \sum_{s=1}^{R-1} P_{i,s} = \frac{1}{1 + \sum_{s=1}^{R-1} e^{\beta_s X_i}}. \quad (14)$$

This constraint can also be introduced by taking $\beta_R = 0$, which produces an identical expression from substitution into expression (13).

For estimates of the betas, the maximum likelihood estimation procedure, described above for the two-class case, is generalized to the R -class case in a straightforward manner.

Proportion Estimation

Although the discussion above has stressed the use of the logit model for classification, it may also be used for proportion estimation. Nelepka

Klamath National Forest Test Site Location



Figure 1. Index map showing location of area modeled in Klamath National Forest.

et al.¹³ and Woodcock et al.¹⁴ have both discussed this problem using underlying assumptions of multivariate normality. The logit model provides an alternative which does not assume multivariate normality and estimates proportions directly. As in classification, either linear or curvilinear models may be selected, and categorical variables may be readily utilized as well. The difference between application of the logit model as a classifier and as a proportion estimator lies in the nature of the calibration data. For the classifier, training observations of x vectors (for pixel i) are each individually labeled with a single class; in the case of proportion estimation, each observation contains the observed proportions of classes and the associated measurement vectors. These proportions constitute weights, and it is easy to show that the likelihood function becomes (as in the two-class case):

$$\ln(L) = w \sum_{i=1}^n g_{i,2} \beta_1 X_{i,1} + w \sum_{i=1}^n g_{i,2} \beta_2 X_{i,2} - w \sum_{i=1}^n \ln(1 + e^{\beta_1 X_{i,1} + \beta_2 X_{i,2}}),$$

where $g_{i,k}$ is the proportion for the i th observation for class k , and w is the observation weight (which, as a constant, is eliminated in differentiation of the fraction).

ORIGINAL PAGE IS OF POOR QUALITY

Application Example

At the present time, the logit-based classifier has not been tested, although a logit model has been used in a forestry remote sensing problem of proportion estimation. This use is summarized in the paragraphs below. For this application, a linear logit model was devised and fitted to forest species compositional data for northern California, predicting the proportion of timber volume for each of five coniferous tree species at each pixel based on registered terrain data quantifying elevation, slope, and aspect. This research utilizes the Video Image Communication and Retrieval (VICAR) system and the Image Based Information System (IBIS) resident at the University of California, Santa Barbara. VICAR/IBIS, developed at the Jet Propulsion Laboratory (JPL) at Pasadena, California, is a job control language which permits the sequential linking and execution of a vast array of Fortran and Assembler routines in a batch environment. In addition to extensive usage of existing VICAR/IBIS routines, new VICAR and non-VICAR software were developed and/or modified as required for this application.

Logit modeling of species proportions used data derived from the Klamath National Forest, located in northern California (Figure 1). Ranging in relief from 500 to 8,000 feet, the Forest includes 2,600 square miles of rugged terrain in the Siskiyou, Scott Bar, and Salmon Mountains. Little of the area is developed beyond management for timber yield, livestock production, and recreation. A wide variety of distinctive vegetative types is present in the area. Forest vegetation includes such coniferous species as noble, red, white, and douglas firs, ponderosa pine, and incense cedar, as well as several oaks, and typical species of chaparral. Thus, the topographic and vegetational characteristics of the area are well differentiated. Within the Klamath National Forest, a study area including most of the Goosenest Range was selected for logit modeling of species composition from terrain features. This area was chosen because calibration data and Landsat images were readily available for it.

Digital Terrain Model

The logit model for this forestry application requires preparation of digital terrain data. These data, obtained from the National Cartographic Information Center, in Reston, Virginia, are derived from processing of 1:250,000 contour maps, and include elevations at every point on a grid of approximately 65 m spacing. Although the data are comparable in scale to a Landsat image, the elevation values are quite generalized because they are produced from small scale contour maps by interpolation.

Slope angle and slope aspect channels can be produced using the elevation data of the registered terrain image. Although a number of slope and aspect generating algorithms are known, the simplest is the fitting of a least squares plane through each pixel and its four nearest neighbors and the calculation of the downslope angle and

direction of the plane. Slope aspect was transformed from a coding of zero to 255 representing 0° to 359° to a cosine function shifted by 45°. This function, proposed by Hartung and Lloyd,¹⁵ contrasts northeast-facing slopes, which present a favorable cool, moist growing environment, with hot, dry southwest-facing slopes. Although the function is defined ecologically, it also simulates Lambertian reflectance from a light source placed in the northeast, and thus the aspect image shown in Figure 2 gives the strong visual impression of relief.

Logit Model

The logit model fitted is

$$\ln\left(\frac{P_k}{P_{-k}}\right) = \beta_{1,k} + \hat{\beta}_{2,k}E + \hat{\beta}_{3,k}A + \beta_{4,k}S, k = 1, 5,$$

where P_k is the probability that a board-foot of timber volume will be drawn from one of five species k , P_{-k} is the probability that the board-foot will not be drawn from species k , E is elevation (compressed to 0-255 range), A is aspect transformed as described above, S is slope angle, and $\beta_{1,k}, \dots, \hat{\beta}_{4,k}$ are the estimated regression constants. Note that five equations, one for each species, actually comprise the model. The model was calibrated using 73 measurements of timber volume prepared by the U. S. Forest Service and located within two subregions of the Goosenest range. These samples are probably not representative of the entire area modeled, but serve for the demonstration purposes of this research. Each sample was located on 1:15,840 scale color air photos and transferred to Band 5 of a registered Landsat image to obtain the line and sample coordinates of the sample point. The coordinates were then used to extract elevation and aspect values for the sample from the registered elevation and aspect images. The coefficients for the model were fitted by a nonlinear optimization algorithm employing the Newton-Raphson method described above.

Given the constants produced by this procedure, the probability images were created using the new VICAR program "PROBMAPS." PROBMAPS, written specifically for this application, calculates the probability of species k for each pixel i using the following expression:

$$P_{i,k} = \frac{e^{\beta_k X_i}}{\sum_{j=1}^5 e^{\beta_j X_i}}, k = 1, 5.$$

PROBMAPS then scales each probability so that the range 0-255 represents 0 to 1. PROBMAPS output images for this example are shown in Figure 2. Brightness values represent that probabilities occurrence for douglas fir, ponderosa pine, white fir, and red fir, with probabilities scales to range from black (0.) to white (1.0). The incense cedar image has been contrast stretched for display purposes, and presents a probability range of 0. to .3 from black to white. The probability images represent maximum likelihood estimates of species proportions; they appear reasonable in light of the known ecological preferences of the species, but their accuracies remain to be determined.

ORIGINAL PAGE IS
OF POOR QUALITY

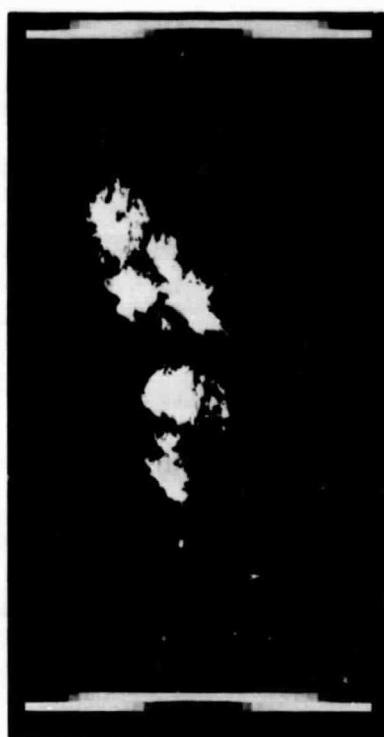


Figure 2. Clockwise from upper left: cosine function of slope aspect; probability images of coniferous forest species red fir; white fir; incense cedar; ponderosa pine; and douglas fir; for Goosenest test area within Klamath National Forest. For probability images, only area within Forest boundary is shown.

**ORIGINAL PAGE IS
OF POOR QUALITY**

Future Work

Although the logit classifier appears to have some unique advantages over conventional maximum likelihood classification; further work will be necessary to prove its value for remote sensing applications. Topics to be investigated include:

1. Model Specification. For what shapes of non-normal distribution are linear models appropriate? Under what conditions are curvilinear models necessary? Could a stepwise procedure, analogous to polynomial curve fitting, be devised for model calibration? Since it is possible to obtain asymptotic estimates of the standard error of each beta, could the stepwise procedure drop individual terms from the model which are not significantly different from zero?
2. Accuracy. How well are probabilities predicted? Can a confidence limit be placed on the predicted probability? Monte Carlo methods may be helpful here. How does accuracy interact with distribution shape?
3. Further Applications. The logit model needs to be exercised on a real classification problem and compared with conventional MLC. Which is more accurate? Which consumes more computational resources? Do categorical variables present any special problems?

These questions and others will be the subject of future research in the application of the logit model to the remote sensing problem.

Table 1. Notation

Term	Definition
p	Number of measurement variables used to characterize each object or observation.
\underline{x}	A p dimensional random vector.
x_i	Vector of measurements on p variables associated with the i th object or observation; $i=1,2,\dots,n$.
$P\{x_i\}$	Probability that a dimensional random vector will take on observed values x_i .
w_k	Member of the k th set of classes w ; $k=1, 2, \dots, K$.
$P\{w_k\}$	Probability that an observation will be a member of class w_k ; prior probability of class w_k .
$P(x w_k)$	Probability density value associated with an observation vector x as evaluated for class w_k .
$\phi_k(\underline{x})$	Probability density value times prior probability for observation vector x evaluated for class w_k .
μ_k	Parametric mean vector associated with the k th class.
Σ_k	Parametric p by p dispersion (variance-covariance) matrix associated with the k th class.

Term	Definition
D_k	p by p dispersion matrix associated with a sample of observations belonging to the k th class; taken as an estimator of Σ_k .
$\sum_{i=1}^n$	Summation sign, add together all occurrences of i from 1 to n .
$\prod_{i=1}^n$	Product sign, multiply together all occurrences of i from 1 to n .
$\hat{\beta}$	Estimated vector of regression parameters.
m_k	Mean vector associated with a sample of observations belonging to the k th class; taken as an estimator of w_k .

References

- ¹Strahler, A.H. (1980). The use of prior probabilities in maximum likelihood classification of remotely sensed data: Remote Sensing of Environment, in press.
- ²Chow, C.K. (1957). An optimum character recognition system using decision functions. IREE Trans. Electron. Computers 6, pp. 247-254.
- ³Sebestyen, G. (1962). Decision-Making Processes in Pattern Recognition, MacMillan, New York.
- ⁴Nilsson, N.S. (1965). Learning Machines - Foundations of Trainable Pattern - Classifying Systems, McGraw-Hill Inc., New York.
- ⁵Schell, J.A. (1973), in Remote Sensing of Earth Resources, Volume I (F. Shahrokh, Ed.), University of Tennessee Space Institute, Tullahoma, TN., pp. 374-394.
- ⁶Reeves, R.G., Anson, A., and Landen, D. (1975). Manual of Remote Sensing, Amer. Soc. of Photogrammetry, Falls Church, VA, 2 vols., 2144 pp.
- ⁷Bronner, W.G., R.M. Haralick, and I. Dinstein (1971). Spectral parameters affecting automated image interpretation using Bayesian probability techniques: Proc. Seventh Int'l. Symp. on Rem. Sens. of Env., pp. 1929-1948.
- ⁸Domenich, T.A. and McFadden, D. (1975). Urban travel demand: a behavioral analysis. Amsterdam, North-Holland.
- ⁹Cox, D.R. (1970). The analysis of binary data. London, Methuen.
- ¹⁰Mantel, N. and Brown, C. (1973). A logistic re-analysis of Ashford and Snowden's data on respiratory symptoms in British coal miners. Biometrics 29, pp. 649-65.
- ¹¹Wrigley, N. (1975). Analyzing multiple alternative dependent variables. Geographical Analysis 7, pp. 187-95.
- ¹²Schmidt, P. and Strauss, R.P. (1975b). The prediction of occupation using multiple logit models. Int. Economic Review 16, pp. 471-86.

ORIGINAL PAGE IS
OF POOR QUALITY

¹³Nalepka, R.F., Horwitz, A.M., Hyde, P.D., Morgenstern, J.P. (1972), Classification of spatially unresolved objects. Manned Spacecraft Center, 4th Am. Earth Resources Program Rev., Vol. 2.

¹⁴Woodcock, C.E., Smith, T.R., Strahler, A.H. (1979), a new model for estimating proportions of land cover within a pixel (abstr.), Machine Processing of Remotely Sensed Data Symposium.

¹⁵Hartung, R.E., Lloyd, W.J. (1969), Influence of aspect on forests of the Clarksville soils in Dent County, Missouri. J. Forestry 67: 178-182.

Acknowledgements

This research has been supported by NASA grant NSG-2377. The authors would like to thank Joseph Scepan and Tara Torburn for the illustrations, and Debbie Heath and Kathy Bresslin for the typing.

A P P E N D I X D

ORIGINAL PAGE IS
OF POOR QUALITY

THE LOGIT CLASSIFIER
A GENERAL MAXIMUM LIKELIHOOD DISCRIMINANT
FOR
REMOTE SENSING APPLICATIONS¹

Paul F. Maynard
Alan H. Strahler

Dept. of Geography
University of California
Santa Barbara, CA, 93106, U.S.A.

ABSTRACT

The logit classifier, a computationally efficient non-parametric pattern classifier, gives higher classification accuracies than the Bayes classifier when the data are not multivariate normal. Bayesian posterior probabilities are algebraically obtained from maximum likelihood parameters fitted to a logarithmic ratio of predicted probabilities. A simple two-class example is:

$$\ln \frac{P_1}{P_2} = \beta_1 X$$

where P_1/P_2 is the observed odds ratio of class w_1 over class w_2 and the β_1 is the vector of calibration constants, estimated by iteratively maximizing the likelihood function of the model given the observed data in the vector of spectral channels X . The estimated posterior probabilities of class membership are algebraically derived as

$$P_1 = \frac{\beta_1 X}{1 + e^{\beta_1 X}}$$

In a Monte Carlo test on simulated nonnormal data, the logit classifier was significantly superior to Bayes, with the accuracy increased up to 34% in one test. In a test with Landsat MSS data, the logit classifier again gave significant increases in accuracy, some as high as 39%.

1. THE BAYES CLASSIFIER

The most commonly used discriminant function in remote sensing applications is the Bayes maximum likelihood classifier (Bayes MLC). It calculates posterior probabilities of class membership based upon parametric estimation of probability density functions for the different classes and awards the classification to the category with the greatest or most "likely" probability. The Bayes MLC is Bayes optimal, i.e., given a zero-one loss function, the costs of misclassification are minimized if the following two conditions are met: (1) the probability density functions are multivariate normal, and (2) the prior probabilities of each class are correctly specified.

¹Presented at the Fifteenth International Symposium on Remote Sensing of Environment, Ann Arbor, MI, May, 1981.

ORIGINAL PAGE IS OF POOR QUALITY

1.1 PROBLEMS WITH BORDER PIXELS

In the traditional supervised classification approach, training sites are selected to be as pure as possible so as to minimize the presence of pixels from extraneous classes. This reduces the intraclass variance, and because it promotes the formation of tightly defined point swarms in measurement space, it reduces class overlap and confusion. However, in the real world, classes are unlikely to be uniform. Even in large agricultural areas, the same crop can have a different spectral reflectance depending upon the underlying soil type or plant moisture content, etc.

The use of pure training sites ignores the spectral information in a significant number of pixels that can be termed border or mixed pixels which should be included in the information class. These pixels occur in the tails of the distributions, and because they are not sampled in the training sites, the pattern for the class is underdispersed with respect to the true information class. This effect produces a difficulty in the classification of mixed pixels. Since the Bayes MLC does not provide for mixed pixels, the implicit assumption is that mixed pixels are to be classified according to the most probable signature match; the components of the signature which are reflected from the less important classes contained within the pixel are thus treated as random noise. The mixed pixel is typically classified by comparing probability densities within the tails of overlapping multivariate normal distributions.

The classification problem in the simple two-class univariate situation is solved by comparing the values of the probability density functions. The classification is awarded to the class with the greatest value (or density) at each X . The simple two-class univariate discrimination problem with equal prior probabilities and equal variances is shown in Figure 1. The Bayes MLC discriminant function for univariate data defines the critical decision point as the intersection point of the two curves. Statisticians will recognize the area to the left of the intersection point and under the curve of u_1 as the area of Type II classification error for the class u_1 on the left and as the area of Type I error for the class u_2 on the right.

If the variance for class u_1 is underestimated and the true density function is given by the expanded curve u_1^* in Figure 2, the area of induced classification error is the difference in area between the true and the underdispersed curves. The true critical decision point is shifted to the left; in this example an observed X greater than 30 but less than 34 that would have been misclassified into class u_1 is now correctly classified into class u_2 .

1.2 EFFECT OF NORMAL APPROXIMATION ON NONNORMAL DATA

At this point, it becomes relevant to consider the bias that is introduced if the probability density function $P[X|u_i]$ for class i is not multivariate normal. In the univariate case (one spectral channel), a probability density function that is skewed to the right will have two effects on its estimated normal curve: the mode will shift to the right and the variance will increase. Visually, what was a skewed curve has been transformed into a normal curve, with the original curve puffed on the left and diminished on the right.

Data from an actual Lanisat MSS test site is used to show the effects of a normal approximation of a skewed distribution. If there are two distributions, with the class on the left approximately normal, and the one on the right severely skewed, Figure 3 is the result. The normal density function estimate $P[X|u_1]$ is shown as curve u_1 in Figure 4. Three areas of induced bias are easily discerned: (1) the left tails, (2) the right tails, and (3) the area between the parametric mean and the nonparametric mode. A pixel whose value is between 70 and 75 that should be classified as u_1 is now misclassified as u_2 .

ORIGINAL PAGE IS
OF POOR QUALITY

Bayes MLC is "robust" regarding minor infractions of normality close to the mean or mean vector if the Mahalanobis distance δ^2 is fairly large ($\delta^2 > 4$; see Zhezhel 1968.) However, as the class mean for w_i moves closer to the class mean of w_j , the probability of error exponentially increases and there are a greater number of potential errors because the densest parts of the distributions are being overlapped.

The special case of Bayes MLC with equal prior probabilities and equal covariance matrices, or Fisher's Linear Discriminant Function (LDF), has been extensively investigated (see Gupta, 1973). Lachenbruch et al. (1973) and Zhezhel (1968) have studied the performance of Fisher's LDF on nonnormal data. Their results indicated that Fisher's LDF was greatly affected by nonnormality. Zhezhel (1968) showed that the maximum error rate is a decreasing function of the Mahalanobis distance δ^2 and tends to 0 as $\delta^2 \rightarrow \infty$. Thus accuracy for nonnormal classes whose δ^2 value is small will be the most severely affected.

1.3 SUITABILITY OF BAYES MLC FOR REMOTELY SENSED DATA

Several authors (Fu et al., 1969; Brooner et al., 1971; Crane et al., 1972) have noted that remotely sensed data were not strictly multivariate normal. Most studies of the distributional nature of remotely sensed data have been with agricultural data. Given that most large sample agricultural remotely sensed data are nonnormal (the Crane et al. (1972) study showed nonnormality for fields averaging over 2,000 data points), it is safe to assume that smaller samples would show even greater departures from normality through the inclusion of random noise. The probability of error for Bayes MLC is higher in two separate situations: incorrectly specified variance and nonnormal distributions. And obviously, the combination of the above two situations will produce even worse results. Since the need to establish pure training site statistics can lead to problems of underestimation of the variance, and in most remote sensing applications the assumption of strict multivariate normality is not valid, Bayes MLC with parametric estimation of the probability density functions is not an optimal classifier for many remote sensing applications.

2. THE LOGIT CLASSIFIER

Most well known predictive statistical techniques are derived from least squares regression. Unfortunately, the least squares technique assumes that the dependent variable is continuously measured. This produces two problems with data that are nominally classified. It can be shown (Domencich and McFadden, 1975) that the predicted values of the response variable in such a model are best interpreted as predicted probabilities. By definition, probabilities are constrained to lie between zero and one. However, the predictions of the least squares model are unbounded and may range from minus infinity to plus infinity. Consequently, the predictions may lie outside the meaningful range of probability and may be inconsistent with their probabilistic nature. The other problem is related to heteroscedasticity and will be dealt with later.

Although there are several statistically acceptable techniques for classifying nominal data such as the probit, logit, and arctan, the logit model has computational advantages since it is a closed (explicit) functional form with convenient curvature properties for numerical optimization.

2.1 THE LOGISTIC TRANSFORMATION

The term "logit" was coined by Berkson (1944) on analogy of Bliss' use (1934) of the term probit, or probability unit, and is a contraction of the phrase "logistic unit." Based on the logistic curve (see Domencich and McFadden, 1975) the dichotomous logistic response law can be defined (Bock, 1970) by

ORIGINAL PAGE IS
OF POOR QUALITY

$$P_{i1} = \frac{e^{\beta_1 X_i}}{e^{\beta_1 X_i} + e^{\beta_2 X_i}} \quad (1)$$

where P_{i1} is the logistically derived posterior probability that pixel i belongs to class w_1 . These posterior probabilities are obtained by taking the logistic transformation,

$$\ln \frac{P_{i1}}{P_{i2}} = \beta_1 X_i \quad \text{or} \quad \ln \frac{P_{i2}}{P_{i1}} = -\beta_2 X_i ,$$

estimating the betas, and substituting these values back into Eq. (1). The logistic transformation enables the value of the predicted dependent variable to range from minus infinity to plus infinity while the values of the predicted probabilities range from zero to one.

It is relatively straightforward to show that in the multinomial case, i.e., more than two classes, the binomial logit generalizes to

$$P_{ik} = \frac{e^{\beta_k X_i}}{\sum_{s=1}^K e^{\beta_s X_i}} \quad k=1,2,\dots,K \quad (2)$$

where P_{ik} is the posterior probability that pixel i belongs to the k th class, given that there are K classes.

The logit model is distribution independent. Proof of this property has been given by Day and Kerridge (1967); they show that the logit model includes an unknown non-negative scalar function of the observed data. This unknown function gives great robustness to the logit model. The only cost is a loss of efficiency should the function be known to belong to a specific distribution, e.g. multivariate normality. In such a case, Bayes MLC will give greater accuracy because it maximizes an unconstrained likelihood function, as compared to the logit which is constrained such that the class probabilities sum to 1 and the estimated class proportions equal the observed class proportions.

2.2 MAXIMUM LIKELIHOOD ESTIMATION OF PARAMETERS

There are two commonly used statistical techniques for estimating regression parameters: ordinary least squares (OLS) and maximum likelihood. Least squares is computationally simpler and the most frequently used. However, even though taking the logistic transformation has solved the problem of interpreting the predicted dependent variables as probabilities, there remain problems with heteroscedasticity because the observed probabilities are based upon a nominally measured dependent variable. Thus, OLS estimated parameters are not BLUE (Best Linear Unbiased Estimators). Consequently, maximum likelihood estimation is the preferred method.

The first step is the formulation of a likelihood function:

$$\Lambda = \prod_{i=1}^N \frac{n_i!}{r_{i1}! \dots r_{iK}!} \prod_{k=1}^K P_{ik}^{r_{ik}}$$

where N is the total number of observations or pixels, the P_{ik} are defined in

ORIGINAL PAGE IS OF POOR QUALITY

Eq. (2), the n_i are the total number of pixels per observation (always 1 in the separate sample remote sensing context), and r_{ik} is equal to 1 if pixel i belongs to class k ; 0 otherwise. After dropping the constant terms and taking the natural logarithm, the log likelihood function is given by

$$\Lambda' = \sum_{i=1}^N \sum_{k=1}^K r_{ik} \ln P_{ik}$$

The second step is maximizing the log likelihood function. Several computational algorithms exist for such a purpose; the Newton-Raphson technique was chosen for this study (see McFadden, 1973). It iteratively converges the vector of first partial derivatives of the log likelihood function to 0, which by definition is a maximum of the function. Deltas or change increments are calculated by inverting the matrix of second partial derivatives (the information matrix) and postmultiplying this times the vector of first partials. The deltas are added to the betas and the first partials and information matrix are recalculated. If the data are not multicollinear, convergence of the first partial derivatives to 0 with 3 decimal places of accuracy occurs in less than 10 iterations.

Maximum likelihood estimation of the conditional logit model can be shown under very general conditions to provide estimators that are asymptotically efficient and normally distributed. These results can be used to construct approximate large sample confidence bounds for the parameters. The large sample (asymptotic) variances and standard errors are obtained from the diagonal terms of the inverted matrix of second partial derivatives.

3. MONTE CARLO TESTS ON SIMULATED DATA

The two models of concern are the Bayes MLC and the linear logit classifier. The following Monte Carlo experimental design was selected. Three major data groups or effects are generated: (1) univariate normal, (2) slightly skewed univariate, and (3) severely skewed univariate. Each group defined a two-way effect between the Mahalanobis distance between categories and the standard deviation within each category. For analytical simplicity, arbitrarily selected Mahalanobis distances of 4, 9 and 16 were crossed with variances of 4, 9, and 16.

Three 3×3 univariate data distributions were required. All were desired to be univariate normal initially, with appropriate algebraic manipulations added *a posteriori* to produce the differences in the mean and standard deviation and the skew effects. Several statistical software packages generate random normal deviates; the International Mathematical and Statistical Library (IMSL, 1979) was chosen because of its easy accessibility and clear documentation.

Given random uniform deviates distributed with a zero mean and unit variance, it is easy to obtain any desired normal distribution by multiplying by the desired standard deviation and adding the desired mean. The slightly skewed distributions were produced by adding the value of one standard deviation (σ) to all the observations that were less than 1 σ from the mean. The severely skewed distributions were produced in a similar fashion. If the value of X was below two standard deviations from the mean, then 6σ was added to it; if the value was between minus one to two standard deviations below the mean then 3σ was added to it.

Each major effect (normal, slightly skewed, and severely skewed) was generated by crossing the three selected Mahalanobis distances with the three selected variances. Consequently, each major effect was composed of nine different distributions or classes. The skewed major effects were created by applying the previously described operations to only those classes with Mahalanobis distances equal to nine. This skewed only the three classes with $S^2 = 9$ and $\sigma^2 = 4, 9, 16$ and left the other classes normal.

ORIGINAL PAGE IS
OF POOR QUALITY

The results of the test with Landsat MSS data are shown in Tables 4 and 5 (training and test sites are separated). Two facts are easily discerned. (1) In all cells except for Bayes' lemon the classification accuracy is unusually high. (2) The accuracy of the logit was always equal or superior to Bayes. The first fact is explained by the accurate location and specification of the training sites. Many Landsat classification efforts do not have access to accurate "ground truth" maps, and this can introduce random error variance. Also, this classification was based on four easily separable classes. The second fact is explained by the shape of the training and test site histograms (please refer to Figures 7 and 8). By comparison with the slightly nonnormal histograms of the simulated data, many of these histograms are truly skewed and significantly non-normal. The two most skewed test classes, urban and lemon, are the classes in which Bayes MLC is significantly inferior to the logit. The greatest increase in classification accuracy occurred in the lemon test sites, an increase of 39%.

5. CONCLUSION

The logit classifier has been shown to be both theoretically and experimentally superior to the Bayes MLC with the simulated data and the Landsat MSS data used in this study. And, as theoretically predicted, the logit classifier was inferior to Bayes MLC with normal data when δ^2 was large.

In order to confidently generalize the above conclusions, further tests are needed on simulated normal data with different transformations and on Landsat MSS data of greater complexity. Given the results produced by this preliminary study, the logit classifier warrants further investigation as a nonparametric alternative to the Bayes MLC.

Another condition besides nonnormality can give rise to biased results for both the logit and Bayes classifiers. This condition is known as a violation of the Independence of Irrelevant Alternatives (IIA) axiom (McFadden, 1973). Briefly, this axiom states:

the ratio of the probabilities of predicting one alternative over another (where both alternatives have a non-zero probability of occurrence) is unaffected by the presence or absence of any additional alternatives in the choice set (Hensher and Johnson, 1981).

The existence of the IIA property introduces the major problem that failure to ensure that all the alternatives are equally distinct may lead to biased predictions. Both the logit and Bayes classification models violate this assumption. Rather than being equally distinct, most classes sought from remotely sensed data are hierarchically distinct, as Swain and Hauska (1977) have observed. As a result, these investigators have developed a Bayesian hierarchical classifier. Likewise, similar work needs to be done with the logit classifier.

Further work on the logit classifier and its applications with remote sensing should include:

- (1) developing a "nested" algorithm to handle violations of the IIA axiom;
- (2) developing the quadratic logit;
- (3) developing capabilities for using prior probabilities;
- (4) converting key portions of the code to Assembler language, and testing the computational costs of the logit and Bayes classifiers;
- (5) testing the logit classifier with a wide range of applications.

6. REFERENCES

- Berkson, J. (1944), "Application of the logistic function to bioassay", Journal of the American Statistical Association 39, pp. 357-365.
Bliss, C.I. (1934), "The method of probits -- a correction", Science 79, pp. 409-410.

ORIGINAL PAGE IS
OF POOR QUALITY

The results of the test with Landsat MSS data are shown in Tables 4 and 5 (training and test sites are separated). Two facts are easily discerned. (1) In all cells except for Bayes' lemon the classification accuracy is unusually high. (2) The accuracy of the logit was always equal or superior to Bayes. The first fact is explained by the accurate location and specification of the training sites. Many Landsat classification efforts do not have access to accurate "ground truth" maps, and this can introduce random error variance. Also, this classification was based on four easily separable classes. The second fact is explained by the shape of the training and test site histograms (please refer to Figures 7 and 8). By comparison with the slightly nonnormal histograms of the simulated data, many of these histograms are truly skewed and significantly non-normal. The two most skewed test classes, urban and lemon, are the classes in which Bayes MLC is significantly inferior to the logit. The greatest increase in classification accuracy occurred in the lemon test sites, an increase of 39%.

5. CONCLUSION

The logit classifier has been shown to be both theoretically and experimentally superior to the Bayes MLC with the simulated data and the Landsat MSS data used in this study. And, as theoretically predicted, the logit classifier was inferior to Bayes MLC with normal data when δ^2 was large.

In order to confidently generalize the above conclusions, further tests are needed on simulated normal data with different transformations and on Landsat MSS data of greater complexity. Given the results produced by this preliminary study, the logit classifier warrants further investigation as a nonparametric alternative to the Bayes MLC.

Another condition besides nonnormality can give rise to biased results for both the logit and Bayes classifiers. This condition is known as a violation of the Independence of Irrelevant Alternatives (IIA) axiom (McFadden, 1973). Briefly, this axiom states:

the ratio of the probabilities of predicting one alternative over another (where both alternatives have a non-zero probability of occurrence) is unaffected by the presence or absence of any additional alternatives in the choice set (Hensher and Johnson, 1981).

The existence of the IIA property introduces the major problem that failure to ensure that all the alternatives are equally distinct may lead to biased predictions. Both the logit and Bayes classification models violate this assumption. Rather than being equally distinct, most classes sought from remotely sensed data are hierarchically distinct, as Swain and Hauska (1977) have observed. As a result, these investigators have developed a Bayesian hierarchical classifier. Likewise, similar work needs to be done with the logit classifier.

Further work on the logit classifier and its applications with remote sensing should include:

- (1) developing a "nested" algorithm to handle violations of the IIA axiom;
- (2) developing the quadratic logit;
- (3) developing capabilities for using prior probabilities;
- (4) converting key portions of the code to Assembler language, and testing the computational costs of the logit and Bayes classifiers;
- (5) testing the logit classifier with a wide range of applications.

6. REFERENCES

- Berkson, J. (1944), "Application of the logistic function to bioassay", Journal of the American Statistical Association 39, pp. 357-365.
- Bliss, C.I. (1934), "The method of probits -- a correction", Science 79, pp. 409-410.

ORIGINAL PAGE IS
OF POOR QUALITY

Bock, R.D. (1970), "Estimating multinomial response relations", in: R. C. Bose, ed., Contributions to Statistics and Probability: Essays in Memory of S.N. Roy. University of North Carolina Press, Chapel Hill, N.C.

Brooner, W.G., R.M. Haralick, and I. Dinstein (1971), "Spectral parameters affecting automated image interpretation using Bayesian probability techniques", Proceedings of the Seventh International Symposium on Remote Sensing of the Environment, pp. 1929-1948.

Crane, R.B., W.A. Malila, and W. Richardson (1972), "Suitability of the normal density assumption for processing multispectral scanner data", IEEE Transactions on Geoscience Electronics 10, pp. 158-165.

Day, N.E. and D.F. Kerridge (1967), "A general maximum likelihood discriminant", Biometrics 23, pp. 313-323.

Domencich, T.A. and D. McFadden (1975), Urban Travel Demand, a Behavioral Analysis. North-Holland, Amsterdam.

Fu, K.A., D.A. Landgrebe, and T.L. Phillips (1969), "Information processing of remotely sensed agricultural data", Proceedings of the IEEE, LVII, 4, pp. 639-653.

Gupta, S.D. (1973), "Theories and methods in classification: a review", in: T. Cacoullos, ed., Discriminant Analysis and Applications. Academic Press, New York.

GRSU, (1976), Distribution and Acreage of Agricultural Land Use in the Oxnard Plain Study Area, University of California, Santa Barbara.

Hensher, D.A. and L.W. Johnson (1981), Applied Discrete-Choice Modelling, Croom Helm, London.

IMSL, (1979), IMSL Library Reference Manual, Edition 7, IMSL, Houston, Texas

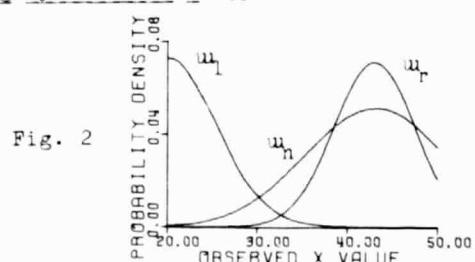
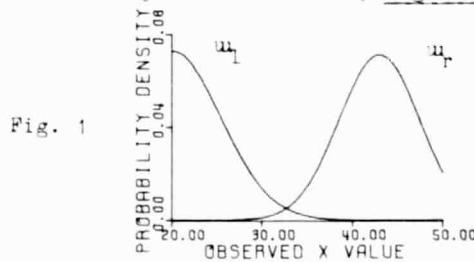
Lachenbruch, P.A. and M.R. Mickey (1968), "Estimation of error rates in discriminant analysis", Technometrics 10, pp. 1-11.

McFadden, D. (1973), "Conditional logit analysis of qualitative choice behavior", in: P. Zarembka, ed., Frontiers in Econometrics. Academic Press, New York.

Swain, P.H. and H. Hauska (1977), "The decision tree classifier: design and potential", IEEE Trans. Geoscience Electronics, GE-15, pp. 142-147.

Wrigley, N. (1976), "An introduction to the use of logit models in geography", Concepts and Techniques in Modern Geography 10, Geo Abstracts Ltd., Norwich, England.

Zhezhel, Y.N. (1968), "The efficiency of a linear discriminant function for arbitrary distributions", Engineering Cybernetics 6, pp. 107-111.



ORIGINAL PAGE IS
OF POOR QUALITY.

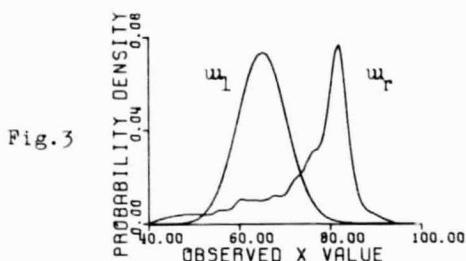


Fig. 3

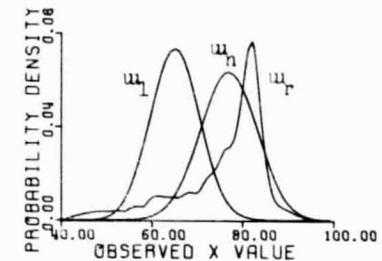


Fig. 4

Figure 5. The nine normal distributions

*	*		*	*		*
*	*	*	*	*	*	*
*	*	*	*	*	***	*
1*	*	1±±1	*	*	*±*±*	±±±1
**	*	±±±±	***	1±±**	±±±±±	
***	*	±±±±	***	±±±±±	±±±±±	1
***	±	±±±±	***	±±±±±	±±±±±	±±±±±
***	1	±±±±	***	±±±±±	±±±±±	1
***	±±	±±±±	***	±±±±±	±±±±±	±±±±±
***	±±±	±±±±	***	±±±±±	±±±±±	±±±±±
***	***	±±±±	***	±±±±±	±±±±±	±±±±±
***	****	±±±±	***	±±±±±	±±±±±	±±±±±
31	*321	±±±±12	*****222	*****1*	*****	*
w ₂	w ₅	w ₇	w ₂	w ₅	w ₇	

Figure 6. The six skewed distributions

Figure 6. The six skewed distributions

Lemon 1 Lemon 2 Barren 1
Barren 2 Veg. 1 Veg. 2

Figure 7. Landsat training sites
(1 is Lansat channel 5, 2 is 7)

Figure 7. Landsat training sites
(1 is Lansat channel 5, 2 is 7)

ORIGINAL PAGE IS
OF POOR QUALITY

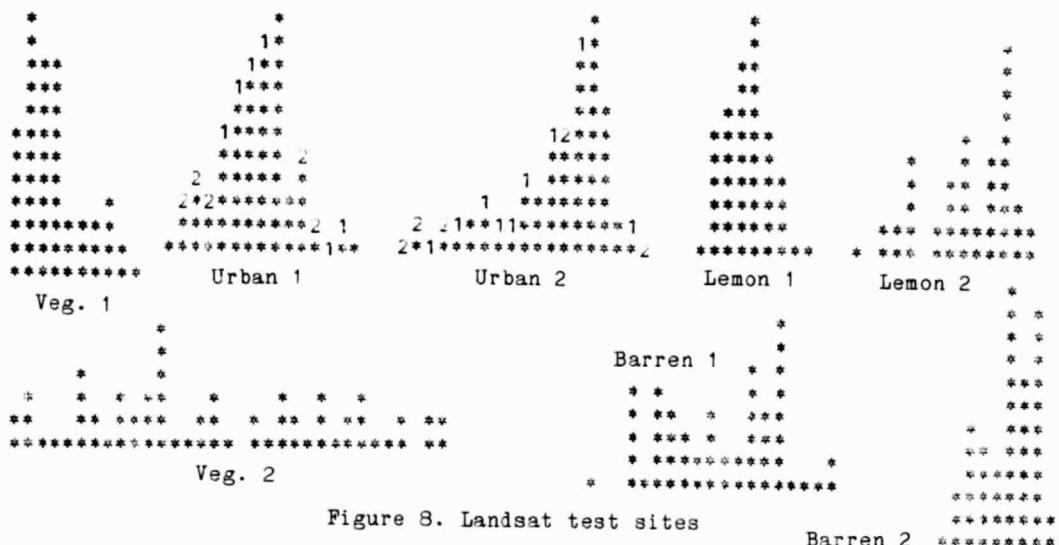


Figure 8. Landsat test sites

Barren 2

δ^2	$\sigma^2=4$		$\sigma^2=9$		$\sigma^2=16$	
	Logit	Bayes	Logit	Bayes	Logit	Bayes
4	.55	.55	.56	.56	.56	.55
9	.39	.39	.28	.28	.40	.37
16	.77	.77	.77	.77	.70	.76

Table 1. Comparison with normal data

δ^2	$\sigma^2=4$		$\sigma^2=9$		$\sigma^2=16$	
	Logit	Bayes	Logit	Bayes	Logit	Bayes
4	.65	.55	.67	.65	.67	.65
9	.66	.51	.64	.56	.60	.53
16	.89	.79	.87	.85	.88	.86

Table 2. Comparison with slightly skewed data

δ^2	$\sigma^2=4$		$\sigma^2=9$		$\sigma^2=16$	
	Logit	Bayes	Logit	Bayes	Logit	Bayes
4	.55	.55	.51	.61	.60	.60
9	.50	.16	.41	.29	.46	.38
16	.81	.77	.81	.81	.84	.82

Table 3. Comparison with severely skewed data

Class	Bayes		Logit		Class	Bayes		Logit	
	Train	Test	Train	Test		Train	Test	Train	Test
Urban	1	20	1	2	Urban	.99	.90	.99	.99
Lemon	0	19	0	1	Lemon	1.0	.59	1.0	.98
Veg.	1	0	1	0	Veg.	.97	1.0	.97	1.0
Barren	0	0	0	0	Barren	1.0	1.0	1.0	1.0

Table 4. Number misclassified

Table 5. Percent correct

A P P E N D I X E

VICAR PROG LOGPR2

ORIGINAL PAGE IS
OF POOR QUALITY

```
//LOGSR JCB (1011,LI,1,2),II,CLASS=D
//FORTRAN PROC
//*
//      FORTRAN EXTENDED ENHANCED          TAS
//FORT   EXEC PGM=IFEAAB,REGION=600K,
//          PARM='OPTIMIZE(3),NAME(MAIN44)'
//STEPLIB DD DSN=SYSLB1.FORTHQ.LOADLIB,DISP=SHR
//SYSPRINT DD SYSOUT=A
//SYSPUNCH DD SYSOUT=B
//SYSTEM  DD SYSOUT=A
//SYSUT1  DD UNIT=SYSDA,SPACE=(TRK,(10,10)),DCB=BLKSIZE=3465
//SYSUT2  DD UNIT=SYSDA,SPACE=(TRK,(10,10)),DCB=BLKSIZE=5048
//SYSLIN  DD DSN=&&LOADSET,UNIT=SYSDA,DISP=(MOD,PASS),
//          SPACE=(CYL,(2,1)),DCB=BLKSIZE=3200
// PEND
//ASSEMBLE PROC
//ASM    EXEC PGM=ASMGASM,REGION=136K,PARM='LOAD,NODECK'
//STEPLIB DD DISP=SHR,DSN=SYSLB1.ASMG.V2L7A.LOADLIB
//SYSLIB  DD DISP=SHR,DSN=SYS1.MACLIB
//          DD DISP=SHR,DSN=UCSB.MACLIB
//SYSUT1  DD DSN=&&SYSUT1,SPACE=(1700,(400,50)),
//          UNIT=(SYSDA,SEP=SYSLIB)
//SYSUT2  DD DSN=&&SYSUT2,SPACE=(1700,(400,50)),UNIT=SYSDA
//SYSUT3  DD DSN=&&SYSUT3,SPACE=(1700,(400,50)),
//          UNIT=(SYSDA,SEP=(SYSUT1,SYSLIB))
//SYSPRINT DD SYSOUT=A
//SYSPUNCH DD SYSOUT=R
//SYSGO   DD DSN=&&LOADSET,DISP=(MOD,PASS),UNIT=SYSDA,
//          SPACE=(80,(200,50))
// PEND
//LNKEDT PROC PREC=SINGLE
//*
//      PLOTTING ADDED 12-80          TAS
//LKED   EXEC PGM=IFWL,PARM=(MAP,LET,LIST),COND=(12,LT),REGION=110K
//SYSLIB  DD DISP=SHR,DSN=WYL.BYS927.STRAHLEE.IPL1.SDSLlib
//          DD DISP=SHR,DSN=SYS1.PLOTLOAD.CALCOMP
//          DD DISP=SHR,DSN=SYS1.PLOTLOAD.UCSB
//          DD DISP=SHR,DSN=SYSLB1.FORTHQ.FORTLIB
//          DD DISP=SHR,DSN=SYSLB1.FORTHEXT.R2L3.FORTLIB
//          DD DISP=SHR,DSN=UCSB.FORTLIB
//          DD DDNAME=&PREC
//DOUBLE  DD DISP=SHR,DSN=SYSLB1.IMSL.LOADLIB.DP8
//SINGLE   DD DISP=SHR,DSN=SYSLB1.IMSL.LOADLIB.SP8
//SYSLMOD  DD DISP=SHR,DSN=WYL.BYS927.STRAHLER.CL'LSLIB
//SYSPRINT DD SYSOUT=A
//SYSUT1  DD UNIT=SYSDA,SPACE=(CYL,(5,1))
//SYSLIN  DD DSN=&&LOADSET,DISP=(OLD,DELETE)
//          DD DDNAME=SYSIN
// PEND
//LOGSR EXEC FORTRAN
C LOGIT, A VICAR PROGRAM THAT USES AN ITERATIVE MAXIMUM LIKELIHOOD
C SOLUTION TO ESTIMATE LOGIT REGRESSION PARAMETERS, THAT CAN BE INPUT
C TO THE VICAR PROGRAM PROBCLAS TO CLASSIFY A DIGITIZED IMAGE.
C
C LOGIT, WRITTEN BY PAUL MAYNARD, GRSU, UCSB, SANTA BARBARA, CA.
C CURRENT VERSION: APRIL, 1981
C
C LOGIT CAN HANDLE UP TO 20 CLASSES, WITH A MAXIMUM OF 30 BETAS.
C THE SECOND JOB REQUIRES 300K REGION.
```

```

COMMON /S1/ DATA(600,20),TDATA(600,20),SDATA(600,20)
COMMON /S2/ VAR(20,50),XM(100),SUBT(40,20),BETA(100)
COMMON /S3/ DELTA(100),FDER(100),ICLASS(600)
INTEGER PNT,ES,EL,SL,SS,SS1,MDEQ
INTEGER*4 PARM(200),QPAR(24),NCLASS(20)
INTEGER*4 KWD(24) /*ITEB*, "ECHO", "APPR", "CBAN", "DBAN", "NCLA",
& "TRAI", "EQUA", "WEIG", "PRIO", "SQUD", "CONV", 12* */ /
INTEGER*4 CLASS/*CLAS*/ , VERT/*VERT*/
REAL TOTAL(600),CONVRG(4)
REAL PRIOR(20),RPARM(200),TMPVCT(100)
LOGICAL BYTE
LOGICAL*1 IBUF(7200)
EQUIVALENCE (IBYTE,BYTE),(PARM(1),RPARM(1))
MAIN PROG DIMENSIONING -- IF CHANGED, CHANGE THESE NUMBERS
NRD=600
NCD=20
NRE=100

```

GET PARAMETERS, AND SET THE DEFAULTS

```

CALL PARM(IND,PARM,200)
NRROWI=PARM(5)
NSAMP=PARM(6)
NPAR=PARM(10)
CALL KSCAN(PARM,NPAR,KWD,QPAR,65000)
NITER=6
IF (QPAR(1).EQ.1) NITER=PARM(KWD(1)+2)
NCOL=0
IF (QPAR(4).EQ.1) NBAND=PARM(KWD(4)+2)
ND=0
IF (QPAR(5).EQ.1) ND =PARM(KWD(5)+2)
NCOLD=NBAND+ND
NS=NSAMP/NCOLD
NCOL=NBAND+1
NCOLD=NCOLD+1
IONE=1
NC=NCOL
NCLS = PARM(KWD(6)+2)
NCLS1= NCLS-1
PNT=KWD(7)+2
IOFF=0
NRROW=0
DO 2 K=1,NCLS
2 PRIOR(K)=1.0
MDEQ=1
IF (KWD(9).NE.0) MDEQ=0
IF (KWD(10).EQ.0) GO TO 6
DO 3 K=1,NCLS
3 PRIOR(K)=RPARM(KWD(10)+K+1)
6 CONTINUE
CONVRG(1)=1.0
CONVRG(2)=0.90
CONVRG(3)=100.
CONVRG(4)=1.0
IF (KWD(12).EQ.0) GO TO 9
NCONV=QPAR(12)
DO 8 I=1,NCONV
8 CONVRG(I)=RPARM(KWD(12)+1+I)
9 CONTINUE

```

ORIGINAL PAGE IS
OF POOR QUALITY.

READ IN IMAGE, ONE LINE AT A TIME

```

C SCAN THROUGH ENTIRE PARAMETER FIELD, AND PLACE DN'S IN
C DATA IF THEY ARE A 'HIT'
CALL OPEN(NLBL,2,0,0,0,32768)
DO 50 IROW=1,NROWI
  IPNT=PNT-1
  IREC=IROW+NLBL
  CALL READ(IND,2,IREC,0,IOFF,NSAMP,IBUF,0)
  CALL CHECK(IND,2)
  IF (IND.EQ.4) GO TO 5010
14  IPNT=IPNT+1
  IF (IPNT.GE.NPAR) GO TO 50
  IF (PARM(IPNT).NE.CLASS) GO TO 15
  IF (PARM(IPNT+2).EQ.VERT) GO TO 40
    IPNT=IPNT+3
    ICLAS=PARM(IPNT-1)
15  CONTINUE
  SL=PARM(IPNT)
  EL=PARM(IPNT+2)+SL-1
  IF (IROW.GE.SL.OE.IROW.LE.EL) GO TO 16
  IPNT=IPNT+3
  GO TO 14
16  CONTINUE
  DO 22 I=SL,EL
    IF (IROW.NE.I) GO TO 22
    SS=PARM(IPNT+1)
    ES=PARM(IPNT+3)+SS-1
    DO 21 J=SS,ES
      NROW=NROW+1
      ICLASS(NROW)=ICLAS
      SS1=J-NS
      DATA(NROW,1)=1.
      DO 20 K=2,NCOLD
        SS1=SS1+NS
        BYTE=IBUF(SS1)
        DATA(NROW,K)=FLOAT(IBYTE)
20    CONTINUE
21    CONTINUE
22    CONTINUE
    IPNT=IPNT+3
    GO TO 14
40  CONTINUE
50  CONTINUE
  IAPPRX=0
  IF (KWD(11).EQ.0) GO TO 113
  NADD=(NCOL*(NCOL-1))/2
  IF (((NCOLD+NADD)*NCLS).LE.NEB.AND.KWD(3).EQ.0) GO TO 63
  WRITE (6,53)
53  FORMAT(/2X,'TOO LARGE DIMENSION FOR COMPLETE QUADRATIC MODEL,
1 APPROXIMATE QUADRATIC MODEL IS USED')
  DO 55 I=1,NCLS
    DO 55 J=1,NADD
55  VAB(I,J)=0.
  DO 59 IR=1,NROW
    IPNT=0
    K=ICLASS(IR)
    DO 59 I=2,NCOL
      DO 59 J=I,NCOL
        IPNT=IPNT+1
59  VAB(K,IPNT)=VAB(K,IPNT)+DATA(IR,I)*DATA(IR,J)
  IAPPRX=1

```

ORIGINAL PAGE IS
OF POOR QUALITY

```
NCOL=NCOL+1
NCOLD=NCOLD+1
IF (ND.EQ.0) GO TO 113
DO 61 IROW=1,NROW
61 DATA(IROW,NCOLD)=DATA(IROW,NCOL)
GO TO 113
63 NEWCOL=NCOL+NADD
IF (ND.EQ.0) GO TO 67
DO 64 IROW=1,NROW
DO 64 I=1,ND
IDUM=I+NCOL
NEWDUM=NEWCOL+I
64 DATA(IROW,NEWDUM)=DATA(IROW,IDUM)
67 DO 103 IROW=1,NROW
NCOL1=NCOL
DO 103 I=2,NCOL
DO 103 J=I,NCOL
NCOL1=NCOL1+1
103 DATA(IROW,NCOL1)=DATA(IROW,I)*DATA(IROW,J)
NCOL=NEWCOL
NCOLD=NEWCOL+ND
113 NR=NROW
IF (NROW.GT.20) NR=20
IF (ND.EQ.0) GO TO 625

TRANSFORM DUMMY VARIABLES INTO 1'S AND 0'S
NCOL1=NCOL+1
K2=0
DO 510 I=1,NROW
DO 510 J=NCOL1,NRB
TDATA(I,J)=0.
510 CONTINUE
DO 600 J=NCOL1,NCOLD
DO 520 I=1,NROW
520 TOTAL(I)=DATA(I,J)
CALL MAXMIN(NROW,TOTAL,XMIN,XMAX,RANGE)
NRANGE=IFIX(RANGE+.1)
IF ((NRANGE+NCOL).GT.NRB) GO TO 605
NVAL=NRANGE+1
DO 580 I=1,NROW
NUM=IFIX(TOTAL(I))
IF (NUM.EQ.0) NUM=NVAL
IF (NUM.EQ.NVAL) GO TO 580
GO TO (531,532,533,534,535),NUM
531 ITCOL=NCOL+1
TDATA(I,ITCOL)=1.
GO TO 580
532 ITCOL=NCOL+2
TDATA(I,ITCOL)=1.
GO TO 580
533 ITCOL=NCOL+3
TDATA(I,ITCOL)=1.
GO TO 580
534 ITCOL=NCOL+4
TDATA(I,ITCOL)=1.
GO TO 580
535 ITCOL=NCOL+5
TDATA(I,ITCOL)=1
580 CONTINUE
NCOL=NCOL+NRANGE
```

ORIGINAL PAGE IS
OF POOR QUALITY

```

600  CONTINUE
      GO TO 610
605  CONTINUE
      IJ=J-1
      WRITE(6,606) IJ
606  FORMAT('OINSUFFICIENT BETA DIMENSIONING TO CONVERT ALL DUMMIES',//,
      & ' ONLY FIRST ',I2,' ARE USED')
610  CONTINUE
      DO 620 I=1,NROW
          DO 620 J=NCOL1,NCOL
              DATA(I,J)=TDATA(I,J)
620  CONTINUE
625  CONTINUE
      DO 819 K=1,NCLS
819  NCLASS(K)=0
C  CALCULATE TOTAL PIXEL NUMBERS IN EACH CLASS
      DO 829 I=1,NROW
          K=ICLASS(I)
829  NCLASS(K)=NCLASS(K)+1
      DO 839 K=1,NCLS
          IF (NCLASS(K).LE.0) WRITE (6,843) K
839  CONTINUE
843  FORMAT (//'* WARNING: CLASS',I3,' HAS NO TRAINING PIXEL,*'
      $, ' YOU ARE GOING TO HAVE TROUBLE'//)
      NB=NCOL*NCLS
      IF (NB.LE.NRB) GO TO 890
          CALL QPRINT('OINSUFFICIENT BETA DIMENSIONING',31)
          GO TO 6000
890  CONTINUE
      DO 900 I=1,NB
          XM(I)=0.
900  EETA(I)=0.
      NB1=NCOL*NCLS1
      DO 903 I=1,NCLS
          IPT=(I-1)*NCOL+1
903  XM(IPT)=1.
      DO 906 IROW=1,NROW
          IPT=(ICLASS(IROW)-1)*NCOL+1
          DO 906 I=2,NCOL
              IPT=IPT+1
906  XM(IPT)=XM(IPT)+DATA(IROW,I)
      DO 913 I=1,NCLS
          IPT=(I-1)*NCOL+1
          DO 913 J=2,NCOL
              IPT=IPT+1
913  XM(IPT)=XM(IPT)/NCLASS(I)
      IF (IAPPRX.EQ.0) GO TO 947
      DO 916 K=1,NCLS
          KPT=(K-1)*NCOL+1
          IVPT=0
          DO 916 J=1,NBAND
              JPT=KPT+J
              DO 916 I=J,NBAND
                  IPT=KPT+I
                  IVPT=IVPT+1
916  VAR(K,IVPT)=VAR(K,IVPT)/NCLASS(K)-XM(IPT)*XM(JPT)
      DO 921 K=1,NCLS
      DO 917 I=1,NADD
917  TMPVCT(I)=VAR(K,I)
      CALL CONVRT(TMPVCT,NADD,NBAND,NCD)

```

ORIGINAL PAGE IS
OF POOR QUALITY

```

DO 918 I=1,NADD
918 VAR(K,I)=TMPVCT(I)/100.
921 CONTINUE
DO 929 K=1,NCLS
DO 923 I=1,NADD
923 VAR(K,I)=VAR(K,I)-VAR(NCLS,I)
929 CONTINUE
DO 99930 K=1,NCLS
KPT=K*NCOL
99930 XM(KPT)=0.
DO 939 IR=1,NROW
DO 939 K=1,NCLS
SDATA(IR,K)=0.
IVPT=0
DO 931 J=2,NC
DO 931 I=J,NC
IVPT=IVPT+1
931 SDATA(IR,K)=SDATA(IR,K)+DATA(IR,J)*DATA(IR,I)*VAR(K,IVPT)
IF (K.NE.ICLASS(IR)) GO TO 939
DATA(IR,NCOL)=SDATA(IR,K)
KNCL=K*NCOL
XM(KNCL)=XM(KNCL)+SDATA(IR,K)
939 CONTINUE
DO 941 K=1,NCLS1
KNCL=K*NCOL
941 XM(KNCL)=XM(KNCL)/NCLASS(K)

ECHO BACK INPUT DATA
947 NR1=NE
IF (KWD(2).NE.0) NR1=NROW
IF (QPAB(2).EQ.1) NR1=PARM(KWD(2)+2)
IF (ND.EQ.0) GO TO 750
WRITE(6,700) NCOL
700 FORMAT('1INPUT DATA, WITH CONTINUOUS CHANNELS IN FIRST ',I2,' COLUMNS
& MN GROUPS',//,' FOLLOWED BY TRANSFORMED CATEGORICAL DATA')
DO 710 I=1,NR1
710 WRITE(6,715) ICLASS(I),(DATA(I,J),J=1,NCOL)
715 FORMAT(' CLASS',I3,2X,20F6.0)
GO TO 800
750 CONTINUE
WRITE(6,770)
770 FORMAT('1INPUT DATA')
DO 780 I=1,NR1
780 WRITE(6,715) ICLASS(I),(DATA(I,J),J=1,NCOL)
IF (NR1.EQ.NROW) GO TO 800
WRITE(6,1361)
1361 FORMAT(' THE LAST OBSERVATION IS:')
WRITE(6,715) ICLASS(NROW),(DATA(NROW,J),J=1,NCOL)
800 CONTINUE
WRITE(6,1362) NROW
1362 FORMAT(' TOTAL NUMBER OF OBSERVATIONS =',I4)

CALCULATE PROPORTION TOTALS
DO 810 K=1,NCLS1
DO 809 J=1,NCOL
SUBT(J,K)=0.
DO 805 I=1,NROW
TK=0.
IF (ICLASS(I).EQ.K) TK=1.
SUBT(J,K)=SUBT(J,K)+DATA(I,J)*TK

```

ORIGINAL PAGE IS
OF POOR QUALITY

```

805      CONTINUE
809      CONTINUE
810      CONTINUE
C
C   CALCULATE DEGREES OF FREEDOM AND NROWS OF OUTPUT PROPORTIONS
NDF=NROW*NCLS1-NCOL*NCLS1
C
C   START OF MAIN LOOP
C
MISPRE=2*NROW
DO 1000 M=1,NITER
WRITE(6,950) M
950 FORMAT('0 ITERATION ',I2)
953 CALL LOGIT(NROW,NCOL,NCLS,MISPRE,
&NRD,NCD,NRB,IER,CONVRG,IAPPRX,NADD,MIS)
IF (IER.EQ.129) GO TO 1060
IF (IER.NE.34) GO TO 911
904 WRITE(6,905)
905 FORMAT(' INVERSE ACCURACY SUCCESSFULLY IMPROVED')
GO TO 1075
911 CONTINUE
C
C   OUTPUT ITERATION VALUES
DO 980 K=1,NB1
BETA(K)=BETA(K)+DELTA(K)
IF (KWD(2).NE.0) WRITE(6,975) BETA(K),FDER(K),DELTA(K)
975 FORMAT(' BETA ',F10.4,' FDER ',F12.4,' DELTA ',2F10.4)
980 CONTINUE
1000 CONTINUE
MISPRE=0
CALL LOGIT(NROW,NCOL,NCLS,MISPRE,
&NRD,NCD,NRB,IER,CONVRG,IAPPRX,NADD,MIS)
C
GO TO 1075
1060 WRITE(6,1061)
1061 FORMAT(' INFORMATION MATRIX SINGULAR',//,' ITERATION HALTED')
MIS=100
1059 DO 1058 I=1,NB1
1058 BETA(I)=BETA(I)-DELTA(I)
1075 CONTINUE
C
C   WRITE OUT RESULTS
WRITE(6,1100)
1100 FORMAT('1FINAL ESTIMATION OF PARAMETERS',//,' THE FIRST SUBSCRIPT
& FOLLOWING THE BETA REFERS TO THE CLASS NUMBER',//,' THE SECOND SUB
& SCRIPT REFERS TO THE PARAMETER NUMBER')
KNT=0
DO 1200 K=1,NCLS1
IF (PRIOR(NCLS).EQ.1.) GO TO 1129
PRIOR(K)=PRIOR(K)/PRIOR(NCLS)
1129 ADJ=NCLASS(NCLS)*PRIOR(K)/NCLASS(K)
ADJ=MDEQ*ALOG(ADJ)
K2=0
DO 1200 J=1,NCOL
K2=K2+1
KNT=KNT+1
IF (IAPPRX.EQ.0.OR.J.NE.NCOL) GO TO 1133
DO 1131 I=1,NADD
VAR(K,I)=VAR(K,I)*BETA(KNT)
TEMP=VAR(K,I)*DELTA(KNT)

```

ORIGINAL PAGE IS
OF POOR QUALITY

```

      JPI=J+I-1
1131  WRITE(6,1150) K,JPI,VAR(K,I),TEMP,XM(KNT),BETA(KNT)
      GO TO 1200
1133  IF (J.EQ.1) BETA(KNT)=BETA(KNT)+ADJ
      WRITE(6,1150) K,K2,BETA(KNT),DELTA(KNT),XM(KNT)
1150  FORMAT(' BETA(',I2,',',I2,')=',F12.6,
      & ' DELTA=',F9.5,' INPUT MEAN=',F10.3,F9.5)
1200  CONTINUE
      WRITE(6,1500) MIS
1500  FORMAT('OPREDICTED PROBABILITIES OF MISCLASS-PIXELS',I5)
      IF (MIS.GE.80) NROW=80
      DO 2000 I=1,NROW
          IF (TDATA(I,ICLASS(I))-GT.0.5) GO TO 2000
          WRITE(6,1600) I,ICLASS(I),(TDATA(I,J2),J2=1,NCLS)
1600  FORMAT(2I4,20F6.2)
2000  CONTINUE
2000  CONTINUE
      CALL END
5000  CALL QPRINT('OPARAMETER ERROR',16)
      GO TO 6000
5010  CALL QPRINT('OUNEXPECTED EOF',14)
6000  CALL QPRINT(' EXECUTION TERMINATED',21)
      CALL ABEND
      RETURN
      END
      SUBROUTINE CONVRT(X,M,N,NR)
      REAL X(M),S(20,20),SINV(20,20),WK(1000)
      IDGT=1
      MPT=0
      DO 100 I=1,N
          DO 100 J=I,N
              MPT=MPT+1
              S(I,J)=X(MPT)
100   S(J,I)=S(I,J)
      CALL LINV2F(S,N,NR,SINV,IDGT,WK,IER)
      MPT=0
      DO 200 I=1,N
          DO 200 J=I,N
              MPT=MPT+1
200   X(MPT)=SINV(I,J)*(2-I/J)
      RETURN
      END
      SUBROUTINE LOGIT(NROW,NCOL,NCLS,MISPRE,
      &NRD,NCD,NRB,IER,CONVRG,IAPPRX,NADD,MIS)
      COMMON /S1/ DATA(600,20),TDATA(600,20),SDATA(600,20)
      COMMON /S2/ VAR(20,50),XM(100),SUBT(40,20),BETA(100)
      COMMON /S3/ DELTA(100),FDER(100),ICLASS(600)
      REAL FDNEW(100),DLTNEW(100),CONVRG(4)

IF DIMENSIONING IS CHANGED IN M/PROG, CHANGE THESE ALSO
      REAL TOUT(20),Q(40,20),WT(20,20),KRON(40,40)
      REAL SS(100,100),WKAREA(12000)
      INTEGER NPNT(100)
      NCLS1=NCLS-1
      NB1=NCOL*NCLS1
      NCKC=NCOL-IAPPRX
      NB=NB1+NCOL
      IONE=1
      DO 5 I=1,NCOL
          DO 5 J=1,NCLS

```

ORIGINAL PAGE IS
OF POOR QUALITY.

ORIGINAL PAGE IS
OF POOR QUALITY

```
5 Q(I,J)=0.  
DO 25 J=1,NB  
  DO 25 J2=1,NB  
    SS(J,J2)=0.  
25 CONTINUE  
  
C CREATE LOGIT QUOTIENT  
APPONE=1.-0.1E-30  
DO 500 I1=1,NROW  
  SUM=0.  
  TEMP=1.  
  MODOVF=0  
  DO 100 K=1,NCLS1  
    OUT=0.  
    K1=K-1  
    NP2=K1*NCOL  
    DO 90 J=1,NCOL  
      DSUM=DATA(I1,J)  
      NP2=NP2+1  
      IF(IAPPRX.EQ.1.AND.J.EQ.NCOL) DSUM=SDATA(I1,K)  
      OUT=OUT+DSUM*BETA(NP2)  
80 90 CONTINUE  
91  OUT=OUT+MODOVF  
IF (OUT.GT.170.) OUT=170.  
IF (OUT.LT.-170.) OUT=-170.  
TOUT(K)=EXP(OUT)  
IF ((7.0E75-SUM).GE.TOUT(K)) GO TO 96  
93  MODOVF=MODOVF-100  
SUM=SUM/2.6881E43  
TEMP=TEMP/2.6881E43  
GO TO 91  
96  SUM=SUM+TOUT(K)  
100 CONTINUE  
TOUT(NCLS)=TEMP  
SUM=SUM+TEMP  
SMALL=SUM*0.1E-30  
DO 150 K=1,NCLS  
  IF (TOUT(K).LT.SMALL) TOUT(K)=SMALL  
  TDATA(I1,K)=TOUT(K)/SUM  
  IF (TDATA(I1,K).GT.APPONE) TDATA(I1,K)=APPONE  
  DO 148 J=1,NCOL  
    DSUM=DATA(I1,J)  
    IF(IAPPRX.EQ.1.AND.J.EQ.NCOL) DSUM=SDATA(I1,K)  
    Q(J,K)=Q(J,K)+TDATA(I1,K)*DSUM  
148 CONTINUE  
150 CONTINUE  
  
C CREATE WEIGHT MATRIX (THE INFORMATION MATRIX)  
DO 200 K=1,NCLS1  
  DO 200 K2=1,NCLS1  
    WT(K,K2)=-TDATA(I1,K)*TDATA(I1,K2)  
    IF (K.EQ.K2) WT(K,K2)=TDATA(I1,K)*(1.-TDATA(I1,K))  
200 CONTINUE  
  
C CREATE KRONECKER MATRIX (THE OTHER COMPONENT OF INFO MATRIX)  
DO 250 I=1,NCKC  
  DO 250 J=1,NCKC  
    KRON(I,J)=DATA(I1,I)*DATA(I1,J)  
250 CONTINUE
```

CREATE 'NROW' LAYERS OF INFORMATION MATRIX
 NRQ=-NCOL
 DO 300 NRW=1,NCLS1
 NRQ=NRQ+NCOL
 NCQ=-NCOL
 DO 300 NCW=1,NCLS1
 NCQ=NCQ+NCOL
 NR=NRQ
 DO 280 NRK=1,NCOL
 NR=NR+1
 NC=NCQ
 DO 280 NCK=1,NCOL
 NC=NC+1
 IF (IAPPRX.NE.1) GO TO 270
 IF (NRK.NE.NCOL.AND.NCK.NE.NCOL) GO TO 270
 IF (NRK-NCK) 261,263,265
 261 TEMP=WT(NRW,NCW)*SDATA(I1,NRK)*SDATA(I1,NCW)
 GO TO 280
 263 TEMP=WT(NRW,NCW)*SDATA(I1,NRW)*SDATA(I1,NCW)
 GO TO 280
 265 TEMP=WT(NRW,NCW)*SDATA(I1,NRW)*DATA(I1,NCK)
 GO TO 280
 270 TEMP=WT(NRW,NCW)*KRON(NRK,NCK)
 280 SS(NR,NC)=SS(NR,NC)+TEMP
 300 CONTINUE

500 CONTINUE

GOODNESS OF FIT TEST
 MIS=0
 DO 11990 I=1,NROW
 IF (TDATA(I,ICLASS(I)).LE.0.5) MIS=MIS+1
 1990 CONTINUE
 WRITE(6,11995) MIS
 11995 FORMAT(' MISCLASSIFIED PIXEL BEFORE ITERATION=',I4)
 IF (MIS.GE.MISPRE) GO TO 91056
 MISPRE=MIS*CONVRG(2)
 GO TO 91000
 1056 IEE=34
 GO TO 2000
 1000 CONTINUE

CALCULATE THE FIRST DERIVATIVE
 K2=0
 DO 651 K=1,NCLS1
 DO 650 J=1,NCOL
 K2=K2+1
 FDER(K2)=SUBT(J,K)-Q(J,K)
 50 CONTINUE
 651 CONTINUE
 NB1=NB1-1
 DO 900 I=1,NB1
 IP1=I+1
 697 M=I
 DO 700 J=IP1,NB1
 IF (ABS(SS(M,I)).LT.ABS(SS(J,I))) M=J
 700 CONTINUE
 IF (M.EQ.I) GO TO 707
 DO 705 J=I,NB1
 TEMP=SS(I,J)

ORIGINAL PAGE IS
OF POOR QUALITY

SS(I,J)=SS(M,J)
705 SS(M,J)=TEMP
TEMP=FDER(I)
FDER(I)=FDER(M)
FDER(M)=TEMP
707 CONTINUE
M=I
DO 709 J=IP1,NB1
IF (ABS(SS(I,M))-LT.ABS(SS(I,J))) M=J
709 CONTINUE
IF (ABS(SS(I,I))-LE.ABS(SS(I,M))*CONVRG(3)) GO TO 800
IF (ABS(SS(I,I))-LE.CONVRG(1)) GO TO 760
IF (ABS(SS(I,I))-LE.ABS(FDER(I)*CONVRG(4))) GO TO 770
FDER(I)=FDER(I)/SS(I,I)
SDEL=SS(I,I)
DO 720 J=I,NB1
720 SS(I,J)=SS(I,J)/SDEL
DO 750 J=IP1,NB1
SDEL=SS(J,I)
DO 740 K=I,NB1
SS(J,K)=SS(J,K)-SS(I,K)*SDEL
740 FDER(J)=FDER(J)-FDER(I)*SDEL
NPNT(I)=1
GO TO 900
760 WRITE(6,761) I
761 FORMAT(I4,' SSII TOO SMALL')
GO TO 800
770 WRITE(6,771) I
771 FORMAT(I4,' DELTA TOO LARGE')
800 NPNT(I)=0
FDER(I)=0.
900 CONTINUE
NPNT(NB1)=0
IF (ABS(SS(NB1,NB1))-LE.ABS(FDER(NB1)*CONVRG(4))) GO TO 920
FDER(NB1)=FDER(NB1)/SS(NB1,NB1)
SS(NB1,NB1)=1.
NPNT(NB1)=1
920 DO 950 I=2,NB1
IF (NPNT(I)-EQ.0) GO TO 950
IM1=I-1
DO 940 J=1,IM1
IF (NPNT(J)-EQ.0) GO TO 940
FDER(J)=FDER(J)-FDER(I)*SS(J,I)
SDEL=SS(J,I)
DO 930 K=I,NB1
IF (NPNT(K)-EQ.0) GO TO 930
SS(J,K)=SS(J,K)-SS(I,K)*SDEL
930 CONTINUE
940 CONTINUE
950 CONTINUE
WRITE(6,1000) (FDER(I),I=1,NB1)
1000 FORMAT(1X,10F10.4)
WRITE(6,1020) (NPNT(I),I=1,NB1)
1020 FORMAT(30I3)
997 DO 999 I=1,NB1
DELTA(I)=0.
IF (NPNT(I)-EQ.0) GO TO 999
DELTA(I)=FDER(I)
999 CONTINUE
2000 CONTINUE

ORIGINAL PAGE IS
OF POOR QUALITY

```
RETURN
END
SUBROUTINE MAXMIN(N,ARRAY,XMIN,XMAX,RANGE)
DIMENSION ARRAY(N)
XMIN=ARRAY(1)
XMAX=ARRAY(1)
DO 100 I=1,N
    IF (XMIN.GT.ARRAY(I)) XMIN=ARRAY(I)
    IF (XMAX.LT.ARRAY(I)) XMAX=ARRAY(I)
100   CONTINUE
    RANGE=XMAX-XMIN
    RETURN
END
/* EXEC LNKEDEDT
INCLUDE SYSLIB(LVMIC)
ENTRY LVMIO
NAME LOGPR2(R)
/* EXEC VICAR
*LOG2 JOB (1011,LI,2),II,CLASS=D
*PASSWORD      Z
*ROUTE SAVE TEMP01  WYL.QCV769.LI.BJE.LOG1
REGION,450K
IND,0X24
,LOGPR2,0X24,*,P1
,P1
BAN,8
TER,4
CLA,10
ONV,30.,1.0,0.0001,0.3
QUD
EIG
BAIN
LASS,1, 7,233,3,10 5,99,5,6 116,186,5,6  51,177,2,5
LASS,2, 91,163,4,5 23,151,4,10
LASS,3, 119,85,2,4 101,81,2,5 123,46,5,2  128,161,4,5
LASS,4, 82,139,2,5 54,91,2,5 97,177,4,2  91,203,1,5
LASS,10, 72,178,2,7 34,136,2,5
LASS,6, 119,139,3,5 24,211,3,5
LASS,7, 56,111,2,5 116,125,3,7
LASS,8, 100,219,1,8 40,111,1,10  98,31,1,8  61,58,2,4
LASS,9, 25,101,2,10 113,111,2,5 6,112,1,5
LASS,5, 4,84,2,5 19,124,2,10 122,203,1,5  72,88,2,5
ND
/VTE2.SYSIN DD *
X24=WYL.QCV769.LI.MSZHAN2
/
```

ORIGINAL PAGE IS
OF POOR QUALITY

A P P E N D I X F

VICAR PROG PRCLSQ

```

//PROBSOR JOB (4530,LI,1,2),II,CLASS=D
//FORTRAN PROC
/*          FORTRAN EXTENDED ENHANCED          TAS
//FORT    EXEC PGM=IFEAAB,REGION=600K,
//           PARM='OPTIMIZE(3),NAME(MAIN44)'
//STEPLIB  DD DSN=SYSLB1.FORTHQ.LOADLIB,DISP=SHE
//SYSPRINT DD SYSOUT=A
//SYSPUNCH DD SYSOUT=B
//SYSTEM   DD SYSOUT=A
//SYSUT1   DD UNIT=SYSDA,SPACE=(TRK,(10,10)),DCB=BLKSIZE=3465
//SYSUT2   DD UNIT=SYSDA,SPACE=(TRK,(10,10)),DCB=BLKSIZE=5048
//SYSLIN   DD DSN=&&LOADSET,UNIT=SYSDA,DISP=(MOD,PASS),
//           SPACE=(CYL,(2,1)),DCB=BLKSIZE=3200
// PEND
//ASSEMBLE PROC
//ASM      EXEC PGM=ASMGASH,REGION=136K,PARM='LOAD,NODECK'
//STEPLIB  DD DISP=SHR,DSN=SYSLB1.ASMG.V2L7A.LOADLIB
//SYSLIB   DD DISP=SHR,DSN=SYS1.MACLIB
//          DD DISP=SHR,DSN=UCSB.MACLIB
//SYSUT1   DD DSN=&&SYSUT1,SPACE=(1700,(400,50)),
//           UNIT=(SYSDA,SEP=SYSLIB)
//SYSUT2   DD DSN=&&SYSUT2,SPACE=(1700,(400,50)),UNIT=SYSDA
//SYSUT3   DD DSN=&&SYSUT3,SPACE=(1700,(400,50)),
//           UNIT=(SYSDA,SEP=(SYSUT1,SYSLIB))
//SYSPRINT DD SYSOUT=A
//SYSPUNCH DD SYSOUT=B
//SYSGO    DD DSN=&&LOADSET,DISP=(MOD,PASS),UNIT=SYSDA,
//           SPACE=(80,(200,50))
// PEND
//LNKEDT  PROC PREC=SINGLE
/*          PLOTTING ADDED 12-80          TAS
//LKED    EXEC PGM=IEWL,PARM=(MAP,LET,LIST),COND=(12,LT),REGION=110K
//SYSLIB   DD DISP=SHR,DSN=WYL.BYS927.STRAHLER.IPL1.SDSLIB
//          DD DISP=SHR,DSN=SYS1.PLOTLOAD.CALCOMP
//          DD DISP=SHR,DSN=SYS1.PLOTLOAD.UCSB
//          DD DISP=SHR,DSN=SYSLB1.FORTHQ.FORTLIB
//          DD DISP=SHR,DSN=SYSLB1.FORTHEXT.R2L3.FORTLIB
//          DD DISP=SHR,DSN=UCSB.FORTLIB
//          DD DDNAME=&PREC
//DOUBLE   DD DISP=SHR,DSN=SYSLB1.IMSL.LOADLIB.DP8
//SINGLE   DD DISP=SHR,DSN=SYSLB1.IMSL.LOADLIB.SP8
//SYSLMOD  DD DISP=SHR,DSN=WYL.BYS927.STRAHLER.CLASSLIB
//SYSPRINT DD SYSOUT=A
//SYSUT1   DD UNIT=SYSDA,SPACE=(CYL,(5,1))
//SYSLIN   DD DSN=&&LOADSET,DISP=(OLD,DELFT)
//          DD DDNAME=SYSIN
// PEND
//PROBSOR EXEC FORTRAN
PROBCLAS, A VICAR PROGRAM THAT TAKES PARAMETERS FROM VICAR PROGRAM
LOGIT, AND GENERATES A CLASSIFIED IMAGE BASED ON CALCULATED
PROBABILITIES.

```

ORIGINAL PAGE IS
OF POOR QUALITY

WRITTEN BY PAUL MAYNARD, GRSU, UCSB, SANTA BARBARA, CA. APRIL 1980

PROBCLAS CAN HANDLE 10 CLASSES AND 20 PARAMETERS PER CLASS

INTEGER*4 PARM(600), QPAR(24)

```
INTEGER*4 KWD(24) /*MSS ',','NCLA','BETA','QUAD',20** */
INTEGER SL,SS
LOGICAL*1 IBUF(7200),OBUF(7200)
LOGICAL BYTE,OVER
REAL*4 RPARAM(600),BETA(500),DATA(50),PROB(10),RBUF(7200),TOUT(10)
EQUIVALENCE (PARM(1),RPARAM(1)),(IWORD,BYTE)
IDIM=50
IONE=1

C
CALL PARAM(IND,PARM,600)
SL=PARM(1)
SS=PARM(2)
NL=PARM(3)
NS=PARM(4)
NROWI=PARM(5)
NSAMPI=PARM(6)
NPAR=PARM(10)

C
CALL KSCAN(PARM,NPAR,KWD,QPAR,&5000)
NBAND=1
IF (QPAR(1).EQ.1) NBAND=PARM(KWD(1)+2)
NSAMP=NSAMPI/NBAND
IF (NS.EQ.NSAMPI) NS=NSAMP
NC=PARM(KWD(2)+2)
NB=NBAND+1
IOFF=0
ICNT=KWD(3)+1
NADD=NB
IF (KWD(4).NE.0) NADD=NB+(NB*NBAND)/2
NBT=NADD*NC

C
DO 10 I=1,NBT
    ICNT=ICNT+1
    BETA(I)=RPARAM(ICNT)
10 CONTINUE

C
CALL OPEN(NLBL,2,0,1,0,32768)
CALL LABELB(IND,NLRO,1,0,0,0,0,NS,OBUF,0,32768)
IREC=NLBL+SL-1
DO 100 IROW=1,NL
    IREC=IREC+1
    CALL READ(IND,2,IREC,0,IOFF,NSAMPI,IBUF,0)
    IF (IND.EQ.4) GO TO 5010
    DO 15 I=1,NSAMPI
        BYTE=IBUF(I)
        REUF(I)=FLOAT(IWORD)
15 CONTINUE
    DATA(1)=1.
    DO 50 J=1,NS
        INC=SS+J-(NSAMP+1)
        DO 20 I=1,NBAND
            I2=I+1
            INC=INC+NSAMP
            DATA(I2)=RBUF(INC)
20 CONTINUE
    DO 30 K=1,NC
        OUT=0.
        K1=K-1
        NSTRT=K1*NADD
        DO 25 J2=1,NB
```

ORIGINAL PAGE IS
OF POOR QUALITY

```

NSTRT=NSTRT+1
OUT=OUT+DATA(J2)*BETA(NSTRT)
25    CONTINUE
      IF (KWD(4).EQ.0) GO TO 26
      DO 23 J2=2,NB
          DO 23 J3=J2,NB
              NSTRT=NSTRT+1
23    OUT=OUT+DATA(J2)*DATA(J3)*BETA(NSTRT)
26    PROB(K)=OUT
30    CONTINUE
      BIG=PROB(1)
      IPNT=1
      DO 35 I=2,NC
          IF (PROB(I).LE.BIG) GO TO 35
          BIG=PROB(I)
          IPNT=I
35    CONTINUE
      IWORD=IPNT
      OBUF(J)=BYTE
C
50    CONTINUE
      CALL WRITE(IND,1,0,0,0,NS,OBUF,1)
      CALL CHECK(IND,1)
      IF (IND.NE.0) GO TO 5020
100   CONTINUE
      CALL END
5000  CALL QPRINT("PARAMETER ERROR",16)
      GO TO 6000
5010  CALL QPRINT("READ ERROR, EOF",16)
      GO TO 6000
5020  CALL QPRINT("WRITE ERROR",12)
6000  CALL QPRINT(" EXECUTION TERMINATED",21)
      CALL ABEND
      RETURN
      END

// EXEC LNKEDT
INCLUDE SYSLIB(LVMIO)
ENTRY LVMIO
NAME PRCLSQ(R)
// EXEC VICAR
**PRB2 JOB (1011,LI),II,CLASS=D
**PASSWORD Z
REGION,300K
RESERVE,1,250,152,*,TEMP
IND,VS3M2S3L
,PRCLSQ,VS3M2S3L,TEMP,(1,1,150,250),P1
,P1
ICLA,10
ISS,8
BETA
11.351099
0.209660
0.312733
-0.029806
0.089002
-0.352661
-0.071456
-0.296937
0.026151
81.273758

```

ORIGINAL PAGE IS
OF POOR QUALITY

-0.040335
-0.251047
0.287979
-0.203972
-0.333009
-0.229950
-0.295022
-0.093382
29.555222
-0.013018
-0.217924
0.217246
-0.271850
-0.215495
0.360924
-0.222184
0.005825
29.442383
-0.272880
0.061678
-0.069569
0.078626
0.053037
0.047201
-0.315095
0.070644
-18.294250
0.259566
0.257642
-0.143739
-0.082690
-0.213324
0.289253
-0.206640
0.150007
-15.007524
0.390971
0.103034
0.363757
0.076100
-0.302809
-0.115772
-0.392119
0.070902
11.187192
0.125121
0.019809
-0.022507
0.311939
-0.106202
-0.077852
-0.426686
0.047489
-37.624924
0.166421
0.010644
0.086536
-0.039247
-0.189180
0.379788

ORIGINAL PAGE IS
OF POOR QUALITY

-0.106250
0.230846
15.097011
0.284601
0.015655
0.037642
0.192711
-0.390045
0.159986
-0.433325
0.019994
0.,0.,0.,0.,0.,0.,0.,0.,0.,0.,0.
E,LIST,TEMP,,(1,221,150,29),(PRINT,ZERCES,NOHI)
E,LIST,TEMP,,(1,200,150,29),(PRINT,ZEROES,NOHI)
E,LIST,TEMP,,(1,159,150,29),(PRINT,ZEROES,NOHI)
//VTR2.SYSIN DD *
VS3M2S3L=WYL.QCV769.LI.MSZHANT
//

ORIGINAL PAGE IS
OF POOR QUALITY

A P P E N D I X G

**University of California
Santa Barbara**

**Soil Loss Prediction in a
Geographic Information System Format**

**A Thesis submitted in partial satisfaction
of the requirements for the degree of**

Master of Arts

in

Geography

by

Michael Alan Spanner

Committee in charge:

Professor Alan H. Strahler, Chairman

Professor John E. Estes

Professor David S. Simonett

March 1982

C-2

ORIGINAL PAGE IS
OF POOR QUALITY

The thesis of Michael A. Spanner
is approved:

A. E. J.

Jessup

John H. Fisher

Committee Chairman

March 1982

ACKNOWLEDGEMENTS

I would like to thank the following people for their help and support during the past few years: Dr. Alan H. Strahler, Dr. John E. Estes and Dr. David S. Simonett, for their assistance, critical comments, and for serving as my thesis committee; Fred Mertz, Larry Tinney, Curtis Woodcock, Earl Hajic and Wayne Hallada, for answering many questions regarding image processing; Tod Streich, for his work on the length of slope routine; Joe Scepan, for his photographic work; Gloria Fletcher, for her work on the graphics and posters; Don Louviere, of the Ventura County Soil Conservation Service, for his technical assistance; Jeanell Trant and Sheri Levine, for their work on the field verification crew; Nathan Gale and Nick Graham, for their assistance with statistical procedures; Elaine Ezra and Doug Stow, for their encouragement; Lisa Cushing, for typing this thesis; all of the staff including: Janna, Anne, Teresa, Meryl, Melinda, Sue K. and Susan P. for their administrative and clerical expertise, as well as brightening up the day; Rob and Raul, for the good times; and to everybody else within the Geography Department and the Geography Remote Sensing Unit who helped in my endeavors.

ABSTRACT

Soil Loss Prediction in a Geographic Information System Format

by

Michael Alan Spanner

Soil loss due to erosion from rainfall was successfully predicted for the Santa Paula 7.5 minute quadrangle, Ventura County, California, utilizing the VICAR/IBIS image processing and geographic information system to simulate the Universal Soil Loss Equation (USLE). This work was part of a NASA funded research project investigating methods of incorporating collateral information in Landsat classification and modelling procedures (NSG-2377). Representing the rainfall, soil erodability, length of slope, slope gradient, crop management and soil loss tolerance coefficients of the USLE were data planes generated from digital Landsat MSS data, USGS Digital Elevation Model topographic data, a digitized NOAA isopluvial map and digitized USDA Soil Conservation Service soil maps. The Pearson product moment correlation coefficient, R, of the soil loss values obtained from the developed geobased model to a sample of manually derived soil losses was .91 after a log transform, significant to the .0001 level. The system accurately targeted soil loss problem areas for subsequent analysis by Soil Conservation Service personnel.

Estimates of accuracy for the intermediate data planes representing rainfall, soil erodability, length of slope, slope gradient, crop management and soil loss tolerance ranged from .81 to 100 percent. These intermediate data planes were accurate and flexible in their representation of environmental information. Incorporating slope and elevation information in the land use/land cover classification considerably improved classification accuracies. Watershed characteristics including slope and length of slope were effectively depicted using the DEM topographic data. Digitized soil maps were a powerful information source, allowing soil erodability and soil loss tolerance to be represented in an image format. Useful statistics from the soil maps, topographic data and land cover were obtained for this watershed. Maps and statistics produced from this soil loss information system will be of great assistance to resource managers dealing with the problem of soil loss in agricultural areas.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
ABSTRACT	iv
Chapter	
1 INTRODUCTION	1
2 BACKGROUND	7
Study Area	7
Surface Wash	10
Soil Loss Models	12
3 SOIL LOSS PREDICTION	15
Universal Soil Loss Equation	15
Rainfall (R)	16
Soil Erodability (K)	17
Length of Slope (L)	18
Slope Gradient (S)	19
Crop Management (C)	19
Conservation Practice (P)	21
Soil Loss Tolerance (T)	22
Soil Loss Information System Models	23
4 GEOBASED DATA	28
Isopluvial Map	28
Soil Maps	28
Digital Elevation Model	30
Landsat MSS	32
5 DIGITAL PROCESSING	34
Generalized Processing Flow	34
Soil Erodability	35
Rainfall	39
Digital Elevation Model	41
Length of Slope	43
Slope Gradient	48
Crop Management	50
Conservation Practice	58
Soil Loss Tolerance	59
Predicted Soil Loss	59

6 ACCURACY AND ANALYSIS	64
Land Use/Land Cover Classification	64
Collateral Data	68
Sensitivity Analysis	74
Analysis of Predicted Soil Loss	76
7 CONCLUSION	79
REFERENCES AND SELECTED BIBLIOGRAPHY	88

LIST OF TABLES

1. Land Use/Land Cover Classes	54
2. C Coefficients	57
3. Classification Accuracy Santa Paula Quadrangle	66
4. Collateral and Predicted Soil Loss Dataset Accuracies	73
5. Sensitivity Analysis	74

LIST OF FIGURES

1. Study area location	3
2. Immediate Watershed for Study Area	4
3. Immature Avocado Orchard	9
4. Generalized Processing Flow	36
5. Soil Erodability Image	40
6. Rainfall Image	42
7. Digital Elevation Model	44
8. Length of Slope Image	47
9. Slope Gradient Image	49
10. Landsat Band 5 Image	51
11. Land Use/Land Cover Classification	56
12. Soil Loss Tolerance Image	60
13. Predicted Soil Loss Image	62
14. Differenced Predicted Soil Loss and Soil Loss Tolerance Image	63

CHAPTER 1

INTRODUCTION

The purpose of this research is to demonstrate the potential of Landsat and collateral information in a geographic information system (GIS) format to simulate the Universal Soil Loss Equation (USLE). The Universal Soil Loss Equation, developed by the United States Department of Agriculture (USDA) Soil Conservation Service (SCS), predicts soil loss due to rainfall in agricultural regions (Wischmeier and Smith, 1965). Soil loss caused by erosion from rainfall is a serious problem in the United States. The Pacific Southwest Inter-Agency Committee (PSIAC) estimated in 1971 that sixty-five percent of our agricultural regions require some form of erosion control. Portions of Ventura County are rated as moderate to severe in erosion hazard by the PSIAC.

Variables of the Universal Soil Loss Equation are coefficients of rainfall, soil erodability, length of slope, slope gradient, crop management and conservation practice. The collection of these variables for the USLE is accomplished by Soil Conservation Service personnel on a per site basis. From tables and charts (USDA Science and Education Administration, 1978), as well as measurements obtained in the field, a prediction of soil loss is calculated for an area in terms of tons of soil loss per acre per year. These

sites range from 1-10 acres depending on the homogeneity of the location. A large area determination of soil loss is not possible, due to the nature of the USLE. The input of the Universal Soil Loss Equation to the Video Image Communication and Retrieval Image Based Information System (VICAR/IBIS) allowed a much more extensive area to be assessed for erosion potential. Each site was represented on a georeferenced grid. Variables from the USLE for the Santa Paula 7.5 minute quadrangle in southern Ventura County (Figure 1) were input to VICAR/IBIS, deriving a prediction of soil loss in tons per acre per year for this mixed agricultural and rangeland location.

Increasing accumulations of sediment are being deposited in Mugu Lagoon, a terminus of drainage for the Santa Paula 7.5 minute quadrangle (Figure 2). The siltation of Mugu Lagoon is due largely to the poor soil conservation practices, sparse grass cover and frequent fire in the upper portions of the watershed. New avocado orchards are being introduced in the canyons extending into the steeper mountainous areas. The sparse ground cover provided by the immature avocados, in conjunction with the steep slopes, creates a potential soil loss problem. Mediterranean annual grasses growing in some of the mountainous areas inadequately protect the soil from rainfall induced erosion. Two fires have occurred in the South Mountain area in the past two years. Removal of the natural vegetative cover by fire greatly increases soil erosion rates. Major

ORIGINAL PAGE IS
OF POOR QUALITY.

3

SANTA PAULA QUADRANGLE STUDY AREA

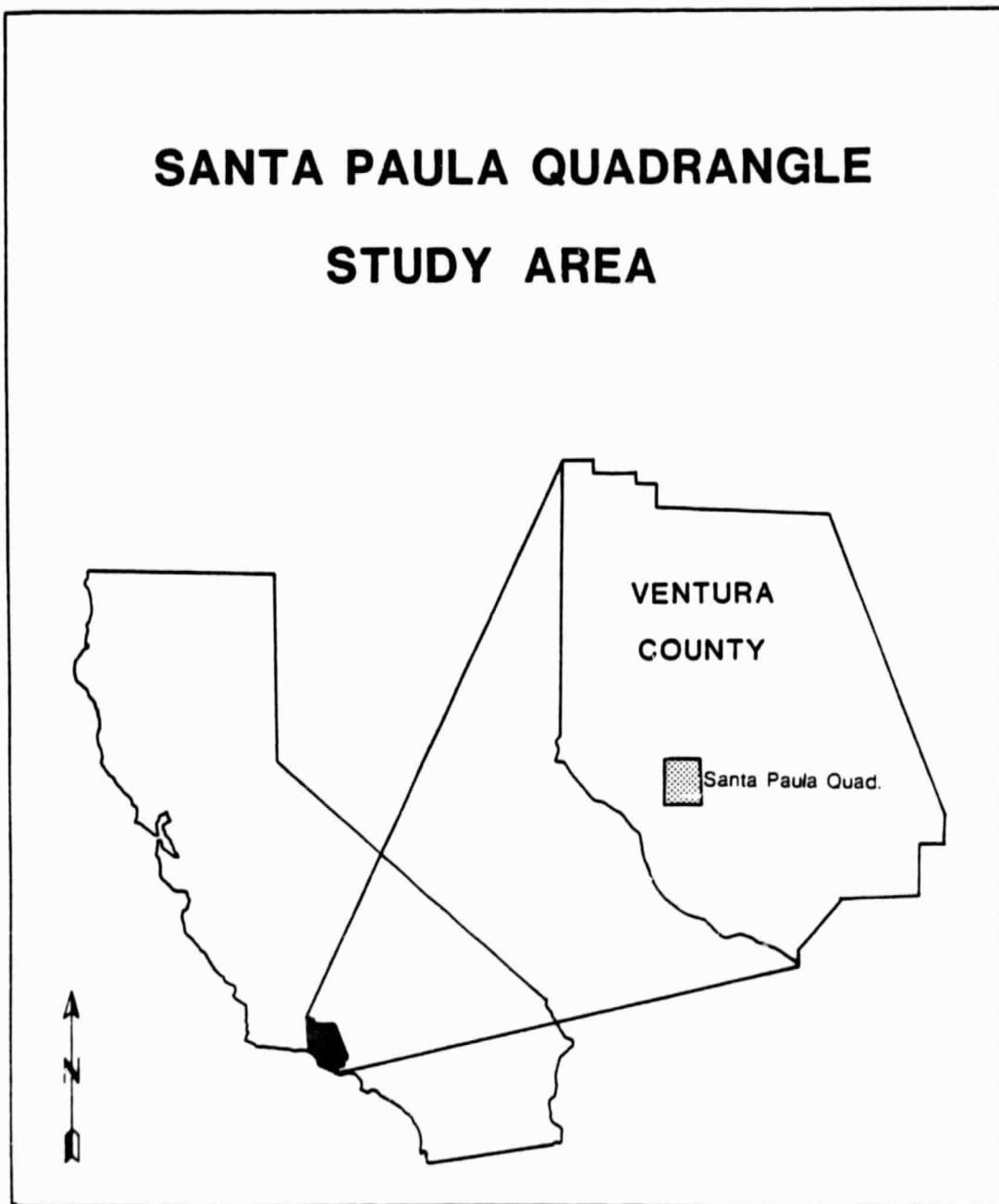


Figure 1 Location of the study area for this Master's Thesis Research, the Santa Paula 7.5 minute quadrangle, Ventura County, California.

ORIGINAL PAGE IS
OF POOR QUALITY

4

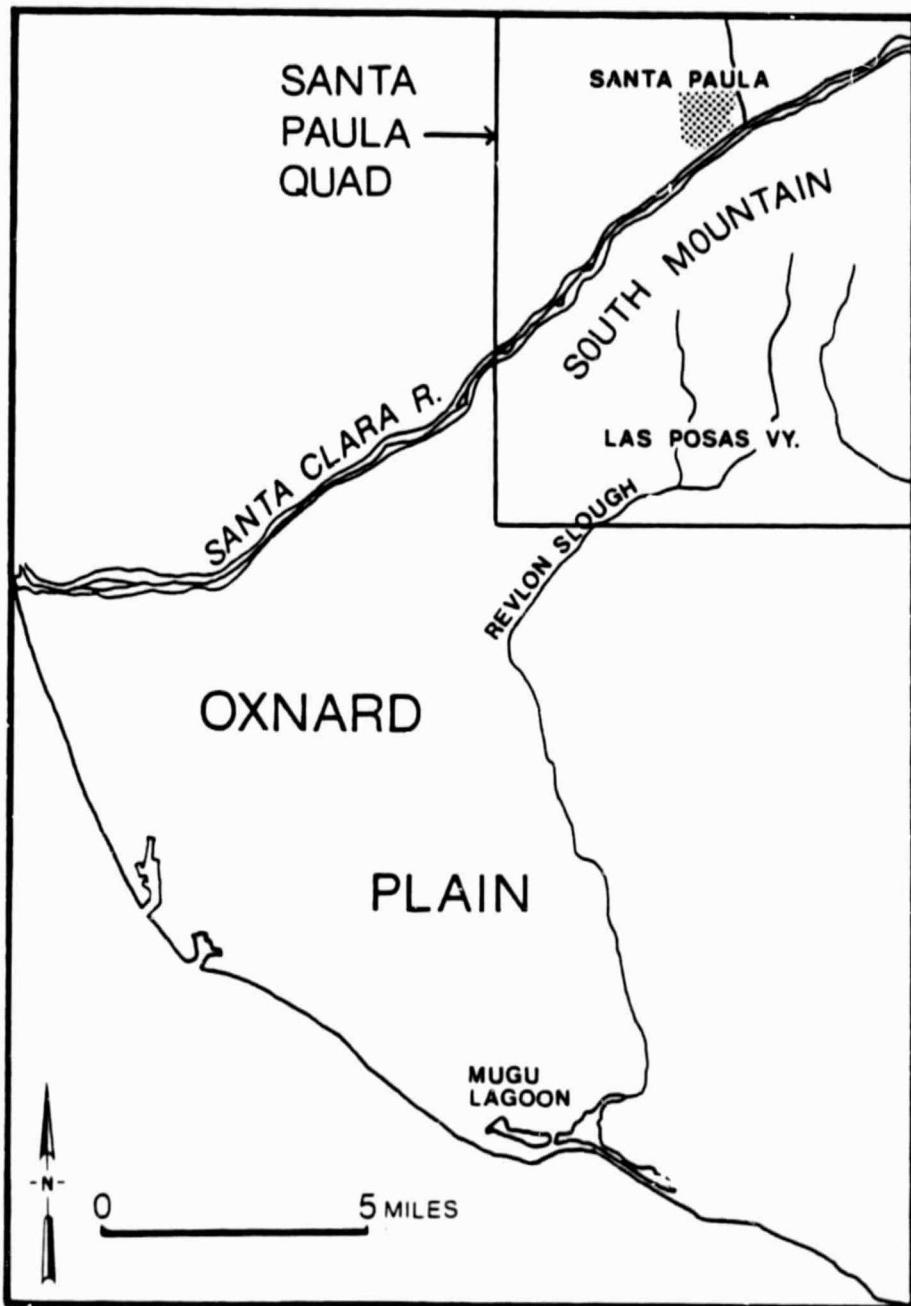


Figure 2 This map shows the location of the major watershed features discussed within the text.

planning and construction efforts are necessary to accomodate the increase in sediment load. A prediction of erosion rates will substantially aid such agencies as the USDA Soil Conservation Service, the U.S. Army Corps of Engineers, the California Coastal Commission, the Ventura County Water Agency and the U.S. Navy facility at Point Mugu for this watershed alone. Implementation of this soil loss model to other locations will assist a wide variety of interests including conservationists, growers, planners, and developers.

Digital image processing for the soil loss study was performed on the Video Image Communication and Retrieval Image Based Information System (VICAR/IBIS). VICAR has been under development at the Jet Propulsion Laboratory (JPL), Pasadena, California, since the mid-sixties. Originally created to process image data from planetary exploration programs, VICAR has been expanded to include applications in earth resources, land use, biomedicine and astronomy. IBIS is a geographic information system which allows the conversion of georeferenced data to an image format for use with remotely sensed data. IBIS was built upon VICAR, permitting image to image registration. Images of different scale from any number of datasets can be superimposed, allowing corresponding pixels to represent the same geographic location (Bryant and Zobrist, 1976).

A land use/land cover classification from digital Landsat data, in conjunction with slope and elevation information from United States Geological Survey (USGS) Digital Elevation Model (DEM)

topographic data, provided the image base for this soil loss study. The slope and length of slope coefficients were also derived from the DEM. USDA Soil Conservation Service soil maps provided the soil erodability and soil loss tolerance information. The rainfall factor was obtained from National Oceanic and Atmospheric Administration (NOAA) isopluvial maps. The soil and isopluvial maps were digitized to facilitate digital processing. The conservation practice factor was not included in this experiment due to the inability of Landsat to resolve conservation practice techniques at a site.

The USDA Soil Conservation Service in Sunis, California provided technical assistance for this work. They also enabled the field verification crew to gain access to several of the ranches in the study area. The Soil Conservation Service is presently pursuing solutions for the problem of soil loss in this watershed. They are concerned with sediment loss reducing crop productivity, as well as the problem of sediment accumulation in the upper watershed and in Mugu Lagoon. A prediction of soil erosion rates will be of considerable assistance to the SCS. As such SCS personnel are interested in the results from this project; this thesis will be presented to them.

CHAPTER 2

BACKGROUND

Study Area

The Santa Paula, California, seven and one-half minute quadrangle comprises the northeastern border of the Oxnard Plain, a fertile region of prime agricultural land (Figure 2). This area displays a typical Mediterranean climate of warm, dry summers and cool, wetter winters. Average annual rainfall is approximately seventeen inches (NOAA Environmental Data and Information Center, 1978). The Oxnard plain is the site of numerous studies by the Geography Remote Sensing Unit, University of California, Santa Barbara (Estes et al. 1979, Stow et al. 1980, Hallada et al. 1981, Tinney et al. 1981). These experiments involved land cover classification, land use/land cover change detection and automatic cluster labelling.

The Santa Clara river, the main drainage for Ventura County runs from northeast to southwest through the Santa Paula quadrangle. The floodplain of the Santa Clara river is the site of many lemon and orange groves, as well as some avocado orchards. These are fairly mature orchards, averaging more than ten years in age. Running east-west through the southern portion of the quadrangle is the

Las Posas Valley. Containing rich alluvial soils eroded from the marine sedimentary sequences of South Mountain, the Las Posas Valley is also the site of lemon and orange groves. Avocados and row crops such as celery and squash are a major source of agricultural income here as well.

The maximum elevation of the Santa Paula quadrangle is approximately 2250 feet on South Mountain, compared to about 200 feet at the lowest point on the Santa Clara River floodplain. New avocado orchards extend into some of the lower canyons of South Mountain and in the foothills north of the Santa Clara River. Avocados are planted on slopes up to and exceeding ten degrees. The steep slopes in conjunction with the sparse cover provided by the immature orchards present a potential soil loss problem in these canyon lands (Figure 3).

Cattle grazing and horse ranching predominate both in the foothills of South Mountain and in the foothills north of the Santa Clara river. Vegetation here consists mainly of Mediterranean annual grasses and chaparral. However, the grass cover is extremely sparse in some locations. There is also a small oak woodland community on the north slope of South Mountain. The hills and mountains are subject to frequent fire. Removal of the natural vegetation due to these fires contributes to accelerated soil losses.

Figure 3 Immature avocado orchard on slope of approximately ten degrees. Orchards such as these are a major cause of increased soil erosion in the Santa Paula area.

Surface Wash

For the purposes of this thesis, discussion of movement of soil downslope is limited to surface wash. Surface wash is defined by Young (1972, p. 62) as:

...the downslope transport of regolith material across the ground surface, through the agency of moving water.

Other forms of movement of debris downslope for the Santa Paula quadrangle include creep, mudflow, earthflow, debris avalanche and debris slide as defined by Varnes (1958). These processes occur in the study area to some extent. However, the dominant downslope form of transport of regolithic materials in the Santa Paula quadrangle is surface wash, as evidenced by field inspection.

There are two distinct processes involved in surface wash. The first is the impact of rain on the ground, termed raindrop impact. The second process, the flow of water across the ground, surface flow, can be subdivided into two categories: sheetwash, whereby the ground is entirely or mostly covered by a moving layer of water; and, rillwash, in which the water flows in very small channels. Channels in rillwash change their size and location on the slope. There are two separate effects of soil detachment by rainfall and runoff: soil detachment, the removal of particles from their original position in the soil; and wash transport, the movement of the detached particles downslope.

Loss of soil in agricultural regions is termed accelerated erosion. Accelerated erosion occurs in locations with cleared vegetation, bare ground or weakened soil structures. This is the case in much of the agricultural portions of the Santa Paula quadrangle.

Under undisturbed conditions, as found in the naturally vegetated regions of the Santa Paula quadrangle, Young (p. 66) discussed three important aspects of soil movement:

- 1) Raindrop impact is a more powerful agent of soil detachment than surface flow, but is relatively less important as an agent of transport. The detachment caused by raindrops does not vary with position on the slope nor, substantially, with slope angle.
- 2) The lowering by surface wash varies directly with the slope angle. The precise relation varies with soil and surface conditions, but a linear proportional increase with the size of the slope angle is the best general approximation.
- 3) The ground loss caused by surface wash varies with approximately the 0.6 power of distance from the slope crest.

In addition, Young (p. 66) made the following discrimination:

In respect of surface wash, a distinction may be made between control by detachment and control by transport. If the agents of transport are more than able to carry away downslope all the material supplied by detachment, the rate of ground loss is determined by the rate of detachment. If control is by wash transport, detached material is supplied faster than it can be transported, so that the transporting media are then always fully loaded.

In other words, under control by wash transport, the rate of soil loss increases lower on the slope, but under control by detachment, the soil loss is independent of distance from the ridge. Slope angle is more important for control by wash transport than control by detachment.

Soil Loss Models

A large number of models have been developed for predicting erosion and sediment yields. With few exceptions these models are based on soil, geologic, climatic, topographic, vegetative and land use/land cover information. Musgrave (1947) developed the first quantitative evaluation of factors in soil erosion. His equation, a first approximation of predicted sheet erosion, is based on soil, crop cover, degree of slope, length of slope and precipitation information. This model is the forerunner of the Universal Soil Loss Equation. The equation has had wide application in the prediction of average annual sheet erosion. A modified Musgrave equation created by substituting the soil erodability and rainfall factors from the USLE has improved this model (Pacific Southwest Inter-Agency Committee, 1968).

The Pacific Southwest Inter-Agency Committee (1968) developed a sediment loss model suitable for use in the Pacific Southwest. A rating system based on surface geology, soils, climate, runoff, topography, ground cover, land use and upland erosion was developed.

This is a highly subjective model; the input variables are overly dependent on user interpretation. Predictions are in terms of low, medium or high sediment loss. To date, this model has not been tested or evaluated in a quantitative manner.

Meeuwig (1971) developed a model predicting soil stability on high elevation rangeland. A predictive equation was developed using soil, vegetation cover, slope, and organic matter content parameters. Meeuwig's equation explains approximately 75 percent of the variance of the log of erosion. The author warned that this equation should not be applied indiscriminately because it was derived from erosion measurements based on fixed amounts of simulated rainfall on small homogeneous plots. The geographic range of this model is therefore limited.

Foster and Meyer (1972) developed a mathematical relationship for the continuity-of-mass transport, and an equation relating detachment of sediment by runoff and sediment load. For known soil, precipitation and topographic characteristics, the erosion pattern along a slope was predicted. This method has sound theoretical development; however, further testing is required to extend the range of its applicability.

A practical model for predicting mass wasting was presented by Swanston et al. (1980) and described by Logan (1981, p. 4), based on:

...a subjective evaluation of the relative stability of an area using soils, geologic, topographic, climatic, and vegetation indicators obtained from aerial photographs, maps and field observations, (and) a limited strength-stress analysis of the unstable sites using available or easily generated field data.

This model was simulated by Logan in a data base approach. However, the Swanston et al. model is still being developed; and, moreover, it is best suited for forest watersheds.

These models are not suitable for soil loss prediction within a geographic information system format for three reasons:

- (1) They are unacceptable for generating image data sets (Meeuwig, 1971; Foster and Meyer, 1972);
- (2) They are not applicable to a mixed agricultural and rangeland region (Meeuwig, 1971; Swanston et al. 1980); and,
- (3) There is insufficient quantitative verification of the model (Musgrave, 1947; Pacific Southwest Inter-Agency Committee, 1968; Meeuwig, 1971; Swanston et al. 1980).

A model that is suitable for soil loss prediction in a geographic information system format, the Universal Soil Loss Equation, is presented in Chapter Three.

CHAPTER 3

SOIL LOSS PREDICTION

Universal Soil Loss Equation

The Universal Soil Loss Equation developed by the United States Department of Agriculture Soil Conservation Service (Wischmeier and Smith, 1965) is appropriate with respect to the three categories discussed in Chapter Two for geographic information system soil loss modelling. The USLE predicts rill and sheet erosion due to rainfall for agricultural regions in terms of tons of soil lost per acre per year. The Universal Soil Loss Equation was originally developed from 10,000 plot years of runoff and soil loss data collected from forty-seven research stations in twenty-four states. Studies were undertaken which measured the contribution of rainfall, soil properties, slope angle, length of slope, crop cover and conservation practice to the loss of soil. A prediction of soil loss was obtained for regions east of the Rocky Mountains. Recent work has extended the USLE to the entire United States, including range and forest lands (USDA Science and Education Administration, 1978).

The basic soil loss equation is:

$$A = R * K * L * S * C * P$$

where:

A = Predicted soil loss in tons/acre/year,
R = Rainfall Factor,
K = Soil Erodability Factor in tons/acre/year,
L = Length of Slope Factor,
S = Slope Gradient Factor,
C = Crop Management Factor, and
P = Conservation Practice Factor.

R,L,S,C and P are dimensionless coefficients. The product of R,K,L,S,C and P provides an estimate of soil loss measured in tons per acre per year.

Rainfall (R)

Research from the Runoff and Soil Loss Data Center of the Agricultural Research Service (ARS), Purdue University, indicates that when factors other than rainfall are held constant, soil loss from cultivated fields is directly proportional to the product of two rainfall characteristics: the total kinetic energy of the storm, and the maximum thirty minute rainfall intensity, (EI) (Wischmeier, 1962). The yearly sum of the EI values from each storm provides the R value, quantifying the impact energy of rainfall in dislodging soil from the ground. The amount of soil which actually leaves the area is affected by the K,L,S,C and P coefficients.

The USDA Soil Conservation Service, Davis, California (1977) has produced a curve relating the two-year, six-hour maximum rainfall to the annual R factor. Two-year, six-hour maximum rainfall

information is available from the NOAA Weather Atlas (Miller et al. 1973) in the form of an isopluvial map. The R value can be determined for all locations in the United States, although the scale of the map is quite small.

Soil Erodability (K)

Soil erodability is defined as the tendency for a specific soil to erode, holding the R,L,S,C and P coefficients constant (Wischmeier and Smith, 1965). The soil erodability, K, is the average soil loss measured in tons per acre per year per unit erosion index, EI, on a unit plot of land. A unit plot is land in continuous fallow, tilled up and down, 72.6 feet long, with a nine percent slope. With the other factors at unity, $K = A/EI$. Several hundreds of plot-years of measurement on twenty-three major soil groups quantified the K value. Based on the rainfall; K was determined for each soil, ranging from .03-.69 (Olson and Wischmeier, 1963).

In a later study (Wischmeier et al. 1971), a soil erodability nomograph was developed which predicts erodability based on five characteristics of the soil. These are percent silt and very fine sand, percent sand greater than .10 mm, organic matter content, structure and permeability. As percent silt and fine sand increase, erodability goes up. By the same token, as the percent sand increases, erodability decreases, an effect produced by the greater

surface area in silt and the greater potential for dislodgement. As organic content increases, erodability decreases, and as the permeability and granularity increase, the erodability is reduced. The twelve K factors in use by the SCS are .10, .15, .17, .20, .24, .28, .32, .37, .43, .49, .55 and .64. The higher the number, the more erodable the soil.

All non-U.S. Forest Service land of Ventura County has been mapped based on soil characteristics by the Ventura County USDA Soil Conservation Service (1969). A project is underway in California to evaluate K factors for all soils. This project has been completed by the SCS personnel for Ventura County; the K values have been calculated for these soils.

Length of Slope (L)

The length of slope factor, L, and the slope gradient factor, S, are calculated separately. However, in the application of the soil loss equation they are evaluated together. Slope length is the distance from a point in a watershed to the origin of runoff for that point, generally a ridge or a peak (Zingg, 1940). The longer the slope, the greater the possibility for erosion due to the increased contribution of runoff from above the site. The potential for partial area flow is increased as well. Length of slope data were collected by Wischmeier et al. (1958) from 136 location years at ten locations. This work indicated that soil loss per unit area

is proportional to some percent of the slope length. The L factor is the ratio of field soil loss to that from one with a 72.6 foot long slope. Therefore, $L = (\lambda/72.6)^{0.3}$, where λ is the length of slope.

Slope Gradient (S)

The slope gradient factor, S, is computed from the nine percent slope unit plot. A nine percent slope has a value of one when the length of slope equals 72.6 feet. An initial study on the effect of slope angle on soil loss by Zingg (1940) concluded that soil loss varies as the 1.4 power of percent slope. Seventeen years of data collected in two locations by Wischmeier et al. (1958) indicated that soil loss is equal to $(.43 + .30S + 0.045^2)$, where S is the percent slope. In California, for slopes less than nine percent, the relation $(0.43 + 0.30S + 0.043S^2) / 6.617$, is used to calculate the S factor. For slopes of nine percent or greater, $(S/9)^{1.3}$ defines the S coefficient (USDA Soil Conservation Service, Davis, 1977).

Crop Management (C)

The crop management factor, C, is the ratio of soil loss from land cropped under the specific conditions encountered, to the theoretical soil loss of the same crop under tilled continuous fallow conditions (Wischmeier, 1960). The C factor is based on the kind

of crop growing in a specific area and the protection it provides against erosion. Crop stages considered in the calculation of the C values are rough fallow, seedling, establishment, growing, mature crop and residue or stubble. Treatment of crop residue has a major impact on the potential for erosion. Residue on the surface provides more protection than residue plowed under. Conditions evaluated in the determination of the C value include: the kind of residues that are left after the crop has been harvested, whether they are removed or remain, and, whether they are left on the surface or plowed under.

Data assessing these variables of crop management were collected from 10,000 plot years of runoff and soil loss data from forty-seven research stations in twenty-four states. Regression and analysis of variance techniques were applied to investigate relationships and interaction effects of crop management on soil erosion. Tables were presented which provide C values under the varying conditions of crop management practice on cultivated agricultural land east of the Rocky Mountains (Wischmeier and Smith, 1965).

Current work has extended the C values to include undisturbed land encountered west of the Rocky Mountains (USDA Science and Education Administration, 1978). Undisturbed land categories include pasture land, rangeland, woodland and idle land. The extension

of C values is based on three separate but interrelated factors:

- (1) Canopy cover. The canopy reduces erosion by rainfall by lessening the impact of raindrops on the soil surface;
- (2) Percent of vegetation in close contact with the ground. 100 percent vegetation cover reduces the impact of rain on the soil to near zero. Vegetation also slows the speed of overland flow, reducing its energy for erosion; and,
- (3) Vegetation litter on the surface. Litter reduces the ability of the rain to erode by lessening the impact of the rain, as well as limiting the transporting ability of runoff.

A chart is provided by the Davis, California, SCS (1977) which displays C values for undisturbed areas based upon canopy cover type and height of cover, percent ground cover and type of ground cover.

Conservation Practice (P)

The conservation practice factor, P, is the ratio of soil loss from the specific conservation practice encountered, to the soil loss from up and down hill cultivation (USDA Science and Education Administration, 1978). Again, this relation is empirically derived from test data in which different types of conservation practice were evaluated. Conservation practices reduce the runoff intensity by altering the direction of the water flow. They also

modify the transport of sediment, generally reducing the effective gradient of the field in question. A value of one is assigned to a plot lacking soil conservation techniques. Practices that reduce soil loss include contouring, contour stripcropping and terracing.

Soil Loss Tolerance (T)

A soil loss tolerance, T, has been calculated; the maximum permissible annual soil loss (USDA Agricultural Research Service, 1961). The soil loss tolerance is quantified in terms of tons of soil loss per acre per year. The T value is assigned by soil scientists based on their judgement of the amount of erosion that will still permit a high level of crop productivity to be sustained economically and indefinitely. A site is considered to have a soil loss problem if the calculated A value exceeds the T value.

Estimates of the T factor are based on the following criteria:

- (1) Maintenance of an adequate rooting depth for crop production.
On shallow soils over hard rock, it is necessary to conserve the soil. A low T value will be assigned;
- (2) Soils that can be renewed with tillage, fertilizers, organic matter and other practices. Renewable soils will be assigned a higher T factor. Nonrenewable soils include soils developed on hard and soft rock with unfavorable nutrient or textural composition. Duripans, natric, strong calcis, gypsic, petrocalsic, and salic soils are included in this category; and,

(3) Value of the nutrients lost and the cost of maintenance of waterways. Ditches may become clogged with sediment; a low T value will be assigned.

The Ventura County SCS has assigned T values for the soil series encountered in Ventura County. Tolerances range from one to five tons per acre per year depending on the field conditions.

Soil Loss Information System Models

Remotely sensed data has been used to derive specific coefficients of the Universal Soil Loss Equation. Morgan et al. (1979) used color and color infrared 70 mm photography at a scale of 1:60,000 to calculate the cropping management factor of the Universal Soil Loss Equation. Detailed information regarding plowing practices and crop residue cover was discernable on both the color and color infrared imagery at this scale. In a later study, Morgan et al. (1980) used the same imagery to calculate the conservation practice factor. They were able to resolve tillage, contour strip-cropping and grass lined waterways for an agricultural region in Wisconsin.

Stephens and Cihlar (1981) analyzed Landsat II, simulated Landsat D (Thematic Mapper), and simulated Spot I data to classify a region composed of pasture, forest and cropland. The simulated Landsat D and Spot I data were resampled to 10 meter resolution. Accuracies were 88 percent, 92 percent and 94 percent for the

Landsat II, Landsat D and Spot I data, respectively. Next, Stephens and Cihlar calculated the correlation between the ratio of the near infrared/red reflectance to the C coefficient for selected study sites. The Pearson product moment correlation coefficients, R, for simulated Spot I, Landsat D and Landsat II data were .97, .95 and .85, respectively. This is an important study because the correlation was determined on the basis of vigor of the vegetation. Further research is necessary along this line of inquiry. The potential of higher resolution sensors with improved spectral characteristics is obvious here.

Researchers have input the Universal Soil Loss Equation into geographic information systems for soil loss prediction. The problem with most of this work is the reliance on manual derivation of topographic data and an inconsistent source of crop management information, or both. Singer et al. (1976) developed a computer simulation of soil loss using the Universal Soil Loss Equation. In this study, vegetation maps provided the crop management factor and topographic data were manually derived from topographic sheets for a rangeland site in northern California. The authors also attempted to model the flow of sediments after erosion based on a computer routine by Kling (1973). However, no basis for the flow of sediments was discussed in this paper.

Degani et al. (1979) presented Soilcart, an interactive computer simulation of soil loss. In this study, crop cover was coded

into the geographic information system from air photographs. Slope and slope length were manually derived from topographic sheets. The interactive aspect of Soilcart make it an important teaching tool; however, for soil erosion modelling, its usefulness is limited.

Patterson and McAdams (1980) input the Universal Soil Loss Equation to a geographic information system to produce erosion hazard potential maps. Landsat MSS data for 1973 and 1978 provided the image base. Slope and length of slope information was derived from soil maps. A high level of flexibility in processing was developed in this study. Land cover for 1973 and for 1978, slope classes, slope length classes, and 1973 and 1978 erosion potential images were displayed. However, topographic data was manually encoded in this work. A truly automated soil loss information system was not developed.

Armond Joyce (1979) applied a modified Musgrave's equation for soil loss prediction in a GIS. The modified Musgrave's equation is a hybrid version of the Universal Soil Loss equation using the soil erodability and rainfall factors from the USLE. This research produced an erosion hazard-forestation needs assessment. Joyce's approach was similar to Patterson and McAdams. The processing for both of these projects was performed at the National Space Technology Laboratory (NSTL) Earth Sciences Laboratory (ERL), Bay St. Louis, Mississippi. The main difference between the two studies was the use of Musgrave's modified soil loss equation and the

manual derivation of topographic factors from topographic maps by Joyce. In addition, a forested watershed was assessed instead of an agricultural region. Again, a fully automated system was not developed in this research.

Berger and Jensen (1980) modeled soil loss and flood potential due to urbanization in a humid subtropical southeastern environment. First, soil loss was predicted using the Universal Soil Loss Equation. Next, the increased rate of runoff was simulated based on a Soil Conservation Service runoff model. Finally, the increased rate of peak flow was predicted using a mathematical model developed at the Georgia Institute of Technology. Large scale, 1:6,000 color and black and white photographs provided land cover information. Topographic data was derived from digital terrain data; however, a methodology for deriving slope and length of slope was not described. Of primary interest is the coupling of two models to derive an assessment of sediment lost and runoff generated. Also unique to this study is the application of the soil loss equation to an urban watershed.

A data base approach for prediction of deforestation induced mass wasting events was accomplished by Logan (1981). A completely automated GIS was used to predict mass wasting in a forested watershed. The erosion model used in this study is a modification of the Swanston et al. (1980) approach. Swanston's model is based on ten natural and management induced hazard factors. Logan reduced

the ten variables to five. These were slope gradient, proximity to roads, soils, precipitation, and vegetation cover. The five factors were transformed into data base images using the VICAR/IBIS image processing system. The images were summed or multiplied based on a design developed by Logan to produce a relative mass wasting index. Landsat MSS data provided the crop management factor and slope information was derived from Digital Terrain Tapes furnished by the National Cartographic Information Center. An automated approach to simulating a mass wasting model in a geographic information system was developed. However, the Swanston et al. model is not appropriate for use in an agricultural location.

A fully automated geographic information system for soil loss prediction in an agricultural region does not exist. Joyce, and Patterson and McAdams approached this ideal with the use of Landsat data, but fell short with their manual derivation of the topographic variables. Berger and Jensen mention the use of digital terrain data to calculate slope and length of slope information, but did not present a methodology. Logan successfully combined Landsat and digital terrain data for a forested watershed. It remains for this thesis to develop an automated geographic information system for soil loss prediction in an agricultural region. The data and processing for an automated soil loss information system are presented in Chapters Four and Five.

CHAPTER 4

GEOBASED DATA

Isopluvial Map

The isopluvial map for the Ventura County area was obtained from the NOAA Precipitation-Frequency Atlas of the Western United States, Southern California subsection (Miller et al. 1973). Two-year, six-hour isopluvials were calculated by NOAA from multiple regression equations based on precipitation measurements from recording and nonrecording gauges. Isopluvials were drawn on maps based on the regression prediction for 47,000 grid points in the western United States. Topography is an important factor in the interpolation of the isopluvials. There are approximately 3300 recording and nonrecording gauges for the western United States. The densest network of gauges are located in the Ventura County area; hence, the isopluvials here are very accurate.

Soil Maps

Soil maps were obtained from the Ventura County office in Somis, California, of the United States Department of Agriculture Soil Conservation Service. The soil maps are at a scale of 1:24,000, and correspond to the United States Geological Survey 7.5 minute topographic sheets. Three soil map sheets comprise a

7.5 minute quadrangle. In the east-west direction each sheet encompasses 7.5 minutes; in the north-south direction they displace 2.5 minutes. Map sheets 27, 23 and 24 arranged north to south represent the Santa Paula 7.5 minute quadrangle. The soil data depicted on these maps were compiled on an orthophoto base created from aerial photographs taken in 1959 and 1965. Cultural features such as roads, railroads and buildings are resolvable on the maps, as are agricultural field patterns. Natural features such as creeks and mountains are distinguishable too.

Soil series and soil phases are delineated on the soil maps. The determination of these soil mapping units is based on characteristics of the soil profile, including the analysis of the horizons as to number, order, thickness, texture, structure, color, organic content and reaction (acid, neutral or alkaline). Soil series have similar horizontal thicknesses and arrangements. Differences in texture, slope and structure of the soil define a soil phase. For example, Calleguas is a soil series, and Calleguas shaly loam, 9-30 percent eroded, is a particular phase of the Calleguas series. For this portion of Ventura County, each soil series includes from one to six soil phases.

When two or more soil series are so intermixed that differentiation between them is difficult or impossible, the resulting association is termed a soil complex. Castaic-Balcom complex 15-30 percent eroded is an example of a complex. In locations where

little, if any, soil development occurs, an alternative designation is used. These soil categories include sandy alluvial plain, landslides, igneous rock land, terrace escarpment, sedimentary rock land, fill land, pits and dumps, riverwash and gullied land. These are not well established soils and the characteristics used to define the other soils do not apply.

Digital Elevation Model

The National Cartographic Information Center under the National Mapping Division distributes Digital Elevation Model topographic data produced by the United States Geological Survey. A DEM is defined by the USGS National Cartographic Information Center (1980, p. 1) as a:

...data file of a sampled array of elevations for a number of ground positions that are at regularly spaced intervals with a defined accuracy.

DEM topographic data for the Santa Paula quadrangle was obtained in the form of a nine track 1600 bit per inch (BPI) computer compatible tape (CCT).

A DEM file may be created from a number of data sources including existing contour plates, profiles of terrain models scanned with stereo photogrammetric equipment, or from computer driven orthophoto devices. The source for the DEM is high altitude photography acquired at a nominal scale of 1:78,000. The data are

processed and resampled to yield one digital elevation model for each 7.5 minute quadrangle. The spatial resolution of the DEM is 30 meters. The vertical resolution is one meter with a root mean square (RMS) elevation error of seven meters. The data are a vast improvement over Defense Mapping Agency (DMA) digital terrain tapes, which are generated from 1:250,000 scale topographic maps and are sampled to a 225 foot horizontal grid. Stow (1978) found that the DMA data failed to represent accurately elevation values due to the coarse resolution of the 1:250,000 topographic sheets. Some values were offset positionally by as many as two pixels.

DEM topographic data are categorized by the USGS National Cartographic Information Center (1980, p. 3) into one of three levels, based on editing, enhancements, and spatial structures.

These levels are:

- (1) DEM 1, a grid of raw elevation data that has been edited for gross blunders and has not been keyed to planimetry;
- (2) DEM 2, elevation data that has been smoothed for consistency, enhanced to remove noise, and filtered to reduce data volume. DEM 2 has not been keyed to planimetry; and,
- (3) DEM 3, elevation data that has been edited and modified to be consistent with planimetric features such as streams, roads and shorelines.

DEM obtained from the NCIC for this study is rated at level one.

DEM topographic data are aligned along Universal Transverse Mercator (UTM) grid lines. The DEM data base is in the initial

stage of development. As of November 1, 1980, fewer than 50 7.5 minute quadrangles were available from the USGS for the state of California.

Landsat MSS Data

Digital Landsat Multispectral Scanner (MSS) data from June 14, 1978 for the Santa Paula area was obtained from the Earth Resources Observation System (EROS) data center, Sioux Falls, South Dakota, in the form of a 1600 BPI nine track computer compatible tape (CCT). The NASA Landsat satellite series consists of three satellites in near polar orbit at an altitude of 920 kilometers. Landsat images the earth using a multispectral scanner (MSS) during its 18 day sun synchronous orbit. A standard Landsat scene encompasses a 185 by 185 kilometer area, containing approximately 7,581,600 picture elements (pixels). The instantaneous field of view (IFOV) is seventy-nine meters, however, due to oversampling in the across track direction, each pixel represents a 79 by 57 meter area on the ground. There are approximately 2340 lines and 3240 samples per Landsat scene. This figure represents an average, due to variation in the sample rate as a result of the variable-velocity scanning mirror on Landsat. The MSS output images in four wavelength increments of the electromagnetic spectrum. These are band 4 .5-.6 microns (green), #5 .6-.7 microns (red),

#6 .7-.8 microns (near infrared), and #7 .8-1.1 microns (near infrared) (Sabins, 1978).

CHAPTER 5

DIGITAL PROCESSING

Generalized Processing Flow

The variables of the Universal Soil Loss Equation were transformed into georeferenced data planes generated from the data sources discussed in Chapter Four utilizing the VICAR/IBIS image processing and geographic information system. Landsat data served as the image base for the soil loss information system. The Landsat data were registered to the Santa Paula 7.5 minute quadrangle. The geometric rectification resampled the Landsat data to sixty meter square pixels (0.9 acre), generating a 231 line by 192 sample image. Sixty meter square pixels, serving as the basic resolution unit for this study, were selected because sixty meters is an even multiple of the thirty meter Digital Elevation Model topographic data.

The rainfall, soil erodability, length of slope, slope gradient and soil loss tolerance images were all registered to the Landsat image base. The rainfall, soil erodability and soil loss tolerance images were generated to correspond to the 60 meter square 231 line by 192 sample grid. DEM, the source for the slope and length of slope images, was resampled to this grid as well. The grid was outlined on the Santa Paula topographic sheet,

facilitating registration and accuracy assessments for this work. The generalized processing flow is depicted in Figure 4.

Soil Erodability

Digitizing routines developed at the Geography Remote Sensing Unit (GRSU), University of California, Santa Barbara, were utilized to create digital data planes from the SCS soil maps on the Talos digitizing table. Three control points were required to transform digitizer coordinates to geographic coordinates. The registration of the maps to the digitizing table created a 231 line by 192 sample sixty meter square grid. Since the 7.5 minute quadrangle is composed of three soil sheets, each map was registered separately.

Each soil mapping boundary was digitized using the Talos electronic cursor, creating files of the X,Y coordinate pairs. Generally, three or four files were necessary to digitize each of the three soil maps. The three or four files for each soil sheet were concatenated into one file for each soil map. The three soil map files corresponding to the Santa Paula 7.5 minute quadrangle were then concatenated into one file. This file was transferred to XYGEN, converting the strings of X,Y number pairs to a file suitable for processing by VICAR/IBIS.

The XYGEN file of line segments was input to the IBIS program POLYSCRB. POLYSCRB transformed the standard polygon file into an image file of polygon borders. Input data in coordinate-point

ORIGINAL PAGE IS
OF POOR QUALITY

36

GENERALIZED PROCESSING FLOW

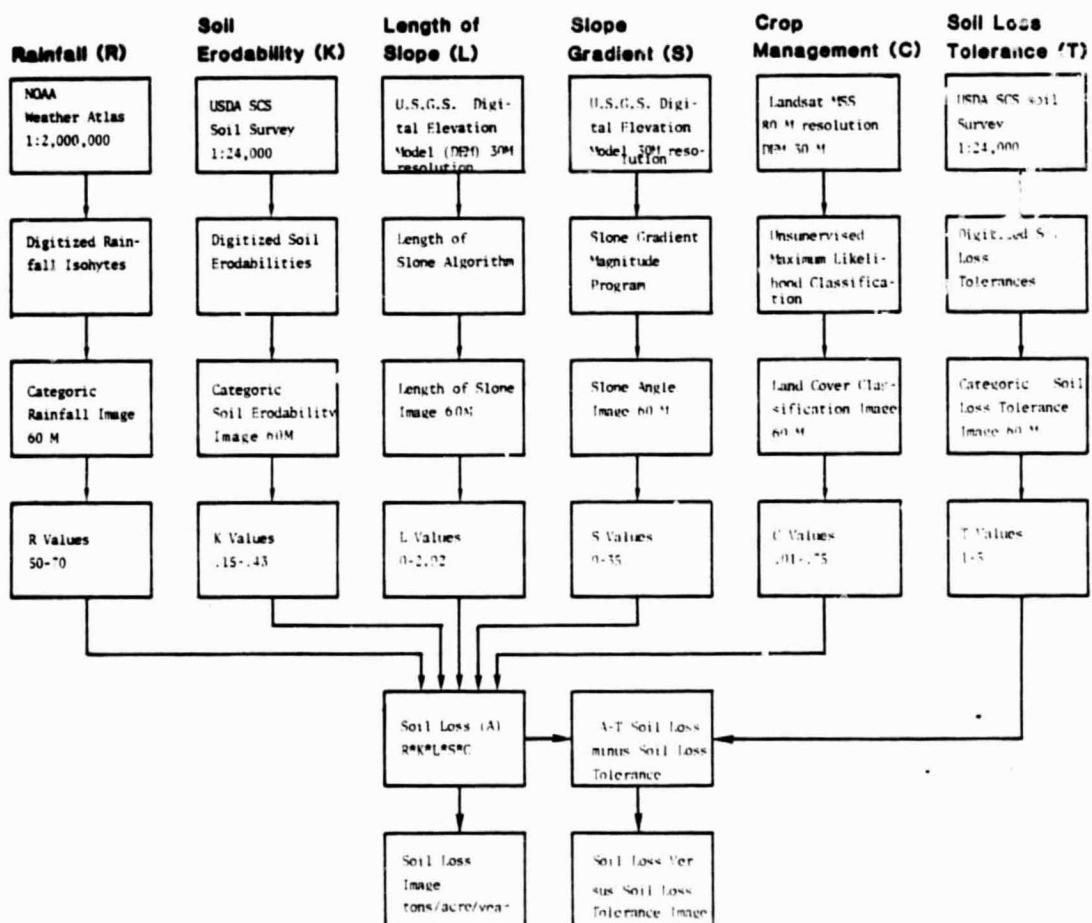


Figure 4 Generalized processing flow for the soil loss information system.

format was used to form a raster base containing polygon outlines. Next, the VICAR program LINEAR was applied to the POLYSCRB'd dataset, transforming the data from byte to halfword, allowing 32,767 increments.

The generation of a halfword dataset was a critical prerequisite for the IBIS program PAINT. PAINT converted the image file of polygon borders from POLYSCRB to an image in which each polygon was represented by a unique DN. An important parameter invoked was the Pborder option. Pborder indicated that the polygon borders on the output file were erased; the borders of the polygons were included as data within the polygon. This was crucial due to the large number of borders on the soil maps. Each border was a pixel wide; the loss of these data would have been unsatisfactory. Unfortunately, the Pborder option of the POLYSCRB routine randomly assigns border pixels to neighboring polygons. This feature reduces the accuracy of the digitized datasets near the borders.

PAINT produced 1498 polygons. These polygons, corresponding to soil mapping units, were reduced to soil erodability classes. On the soil map, the soil erodability for each soil mapping unit was coded, ranging from 1-20. The numbers did not refer to the erodability, but were simply values in a lookup table constructed for ease in data interpretation. Gullied land, landslides, igneous rock land, sedimentary rock land, fill land, riverwash, badland, sandy alluvial plain, terrace escarpment and pits and dumps soil

mapping units were assigned unique numbers, as were the regular soil series and phases.

A labelling procedure was developed to convert each soil polygon to a soil erodability class. The spatial location of each PAINT'ed polygon was matched to the soil mapping unit on the SCS soil maps. The lookup value for each soil mapping unit on the SCS maps corresponded to the number of the polygon generated by PAINT. The VICAR program HSTRETCH was run on this data to reduce the original 1498 polygons to 20 classes.

Reduced K soil polygons ranging from one to twenty were then assigned their appropriate K value based on the Universal Soil Loss Equation. The regular soil series and phases were stretched to .15, .20, .23, .32, .37, and .43. VICAR can process only integer values, therefore, each value was multiplied by 100 before input to the HSTRETCH program. This increase by two orders of magnitude was noted and accounted for later in the processing. In all subsequent descriptions, decimal coefficients will be described. The values were actually converted to integers through a multiplication by one or two orders of magnitude, and then rectified later in the processing.

K values are not calculated by the SCS for the special soil classes because a wider range of erodability within each of these special classes necessitates field examination for each site. It was not possible to examine all of these sites, therefore, they

were deleted from further study; the K values were set to zero. In the final erosion prediction soil losses were zero for the special soil classes, comprising 8.8 percent of the study area. A large majority of these soils are located within the Santa Clara River channel, an unimportant agricultural site. The river class was also excluded from further study, as will be discussed later in the chapter. A K value was assigned to each cell on the soil erodability image, ranging from 0-.43 (Figure 5).

Rainfall

Isopluvials from the NOAA Precipitation-Frequency Atlas (Miller et al. 1973) were transferred to the Santa Paula 7.5 minute quadrangle. Two-year, six-hour precipitation ranged from 1.6 inches in the southern portion of the quadrangle to 2.0 inches in the north. The VICAR/IBIS processing flow for the rainfall coefficient was similar to the processing flow for the soil erodability coefficient. On the Talos digitizing tablet the isopluvials were digitized. The digitizer program linked the X,Y coordinate pairs to XYGEN, converting the X,Y coordinate pairs to a file suitable for further IBIS processing. POLYSCRB was then used to transform the XYGEN isopluvial vector file into an image file containing polygon borders. The raster dataset was input to PAINT, converting the raster file of isopluvial borders to an image with a unique DN assigned to each polygon. The VICAR program HSTRETCH was run on

ORIGINAL PAGE IS
OF POOR QUALITY

40



Figure 5 Registered soil erodability image generated from digitized Soil Conservation Service soil maps. Darker tones represent less erodible soils; lighter tones represent more erodible soils.

the PAINT'ed image, transforming the polygon value to the appropriate R coefficient for each rainfall region. An R coefficient ranging from .50-.70 was assigned to each cell on the grid (Figure 6).

Digital Elevation Model

DEMVIC, a DEM login program, converted the raw DEM data to an unlabeled VICAR image in 16 bit format. The VICAR program VSAR was applied to the output from DEMVIC to convert the file to VICAR format. North was skewed 90 degrees to the east on the DEM. The VICAR program FLOT rotated the data counterclockwise, to orient north to the top of the grid.

The DEM was listed off on a lineprinter for visual inspection. Canyons extending off the boundary of the Santa Paula quadrangle were good features for comparing the DEM to the topographic sheet. The location of points on the DEM was in excellent correspondence with the topographic map. The four corners of the Santa Paula DEM and the Santa Paula 7.5 minute quadrangle were used as ground control points (GCP's) in the registration of the DEM to the quadrangle. The VICAR program GEOMA was utilized for registration.

Two methods to geometrically rectify an image using GEOMA are available through VICAR at UCSB. The first is the nearest neighbor approach, which does not interpolate grey values. The grey value of the nearest pixel from the old grid system is input to the new

ORIGINAL PAGE IS
OF POOR QUALITY.

42

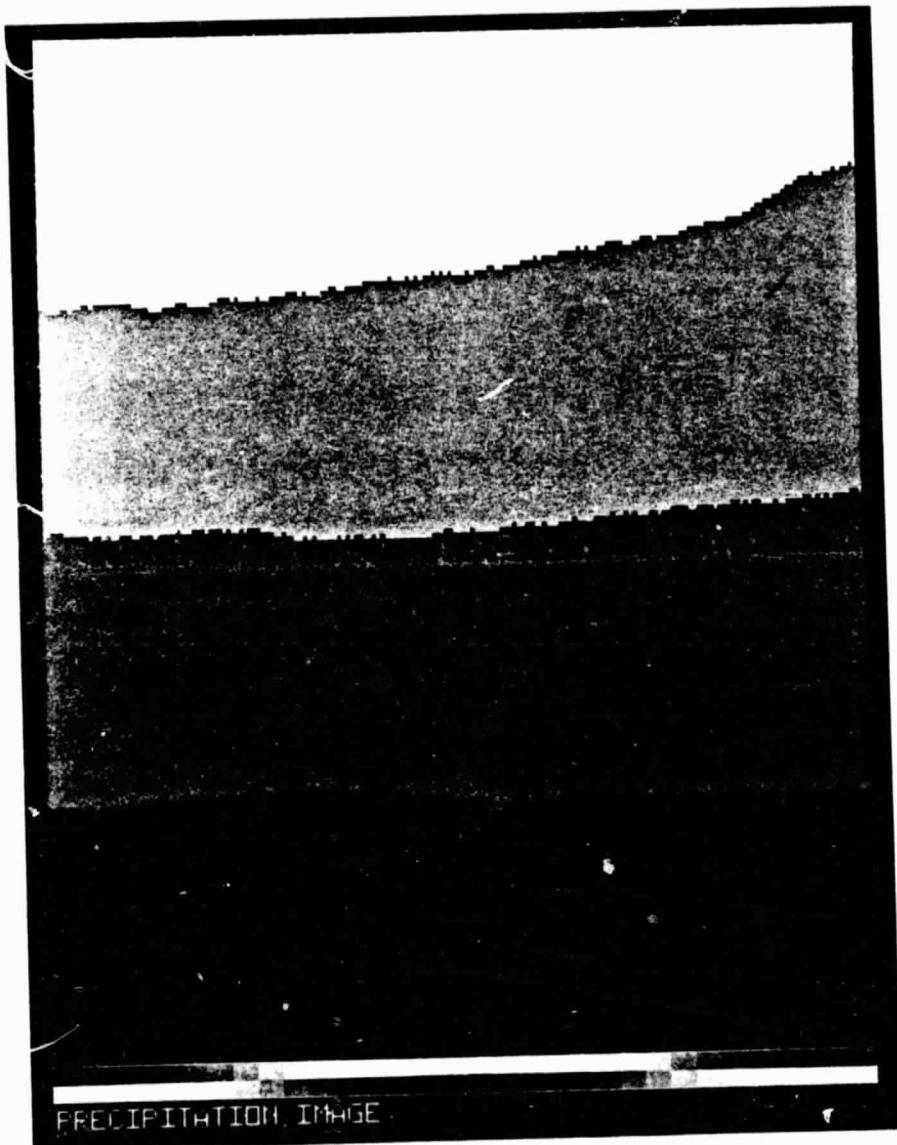


Figure 6 Registered precipitation image generated from digitized NOAA Precipitation-Frequency Atlas. The darkest tone represents two-year, six-hour rainfall of 1.6 inches; the lightest tone represents two-year, six-hour rainfall of 2.0 inches.

grid system. Intensity values do not change; however, the spatial arrangement of the data is altered somewhat. The second method applies a bilinear interpolation function to the data. Here, the new values have a higher spatial correlation to the old data, but the intensity values are altered by the interpolation process (Castleman, 1979). The nearest neighbor approach was used to keep the data as faithful to the original DEM as possible.

Geometric rectification resampled the DEM to a 60 meter square grid of 231 lines by 192 samples to maintain compatibility with the Landsat MSS data. The DEM were tested against the topographic sheet to determine the accuracy of the registration. Approximately 40 points from the grid were analyzed. Peaks of mountains were the best topographic features to test because they were easily located on the topographic sheet and the DEM. Based on the topographic grid, pixels from the DEM corresponded quite well with the topographic sheet. No values were more than one pixel off; most were in perfect correspondence. The DEM elevation image is presented in Figure 7.

Length of Slope

An algorithm that computed length of slope from Digital Elevation Model (DEM) topographic data was developed. Length of slope is the distance from a point in a watershed, to the source of runoff for that point, generally a ridge or a hilltop. A search was conducted

ORIGINAL PAGE IS
OF POOR QUALITY

44

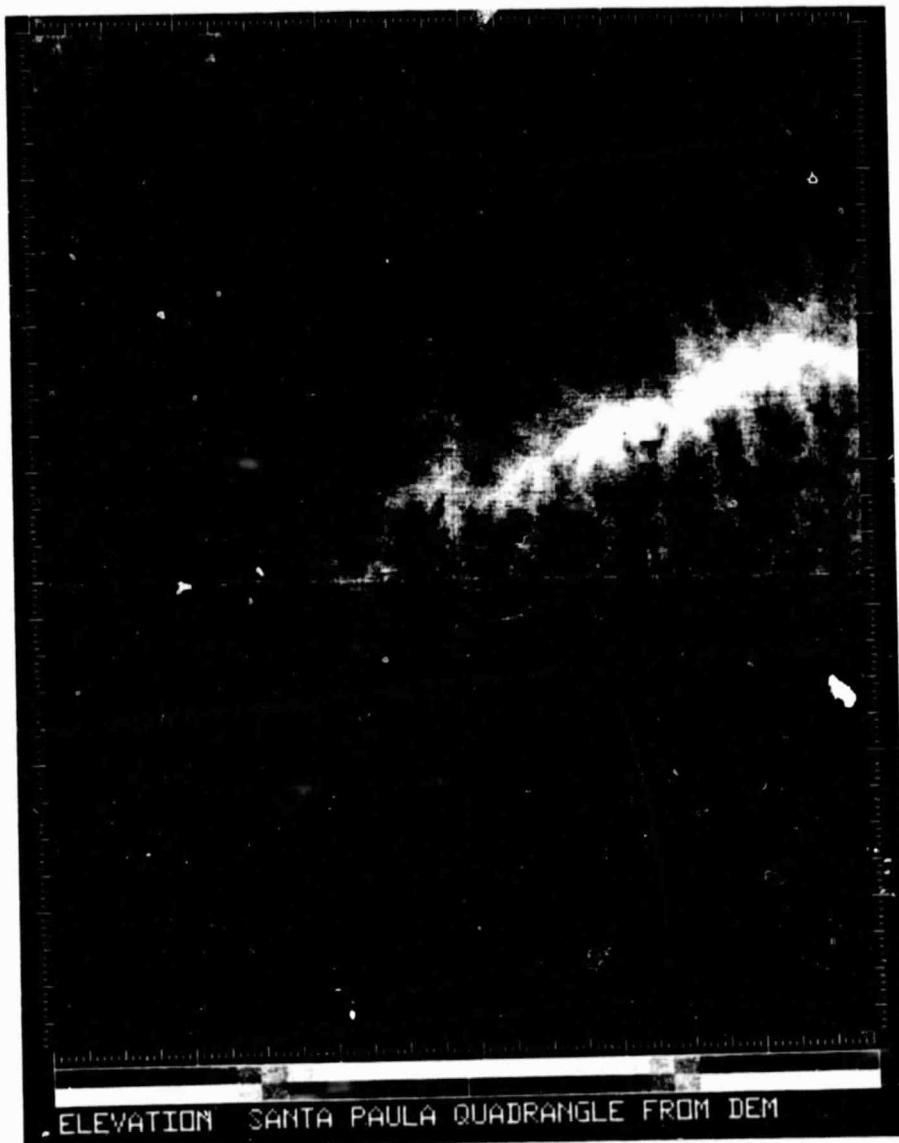


Figure 7 Registered elevation image for the Santa Paula quadrangle from USGS Digital Elevation Model topographic data. Dark tones represent the lowest elevations; light tones represent the highest elevations.

to determine if this calculation using digital topographic data had been developed elsewhere. Collins (1975) developed a package of programs which calculate watershed characteristics from DEM. Length of slope is not included in this package. Junkin (1979), NASA National Space Technology Laboratories (NSTL), Mississippi, developed a set of routines that compute slope, slope length and aspect from Digital Elevation Model topographic data. However, length of slope here is the distance across an elevation cell, a within cell slope length. The sum of cell lengths along a slope to the source of runoff is required for the USLE slope length coefficient. Therefore, it was necessary to develop a length of slope program in-house.

A buffer one pixel wide was applied to the border of the DEM, enabling slope length to be calculated for cells on the margin of the DEM. Starting at line one, sample one, a three by three moving window surrounding this coordinate was created. The direction of the steepest slope between the middle cell of the window and any of the surrounding eight elevation cells was determined. The distance between the middle cell and highest cell surrounding it was calculated by the Pythagorean theorem. The direction of the movement was noted.

The window then moved so that the steepest cell from the old center became the new center. Direction of steepest slope was again determined using the new cell as the center. This direction was

limited to forty-five degrees from the previous movement.

Pythagorean distance was calculated for the new movement. The process continued until the slope did not continue uphill. This was the source of runoff for the initial cell; the routine located a ridge or a peak. The sum of the distance of all the moves was tallied and applied to the initial cell. The window then moved to the next sample same row and repeated the process. The routine continued until the length of slope for all cells of the DEM were calculated.

An important feature of the algorithm is a preview subroutine. If there were two or more equally steep slopes, Preview was called. The preview subroutine set up another window around the equally steep slopes. The steepest slope from these equally steep slopes was determined. The cell with the steepest new slope became the cell for the slope length calculation. Processing then continued as before.

Slope length in meters for each cell was generated by the length of slope routine. This was converted to feet to correspond to the USLE L factor formula (Figure 8). The formula, $L = (\lambda/72.6)^{0.3}$, where λ equals the slope length in feet, was input to the VICAR program F2. F2 allows general arithmetic functions to be applied to one or two input images in byte or halfword format. A halfword slope length coefficient for each cell was produced.

ORIGINAL PAGE IS
OF POOR QUALITY

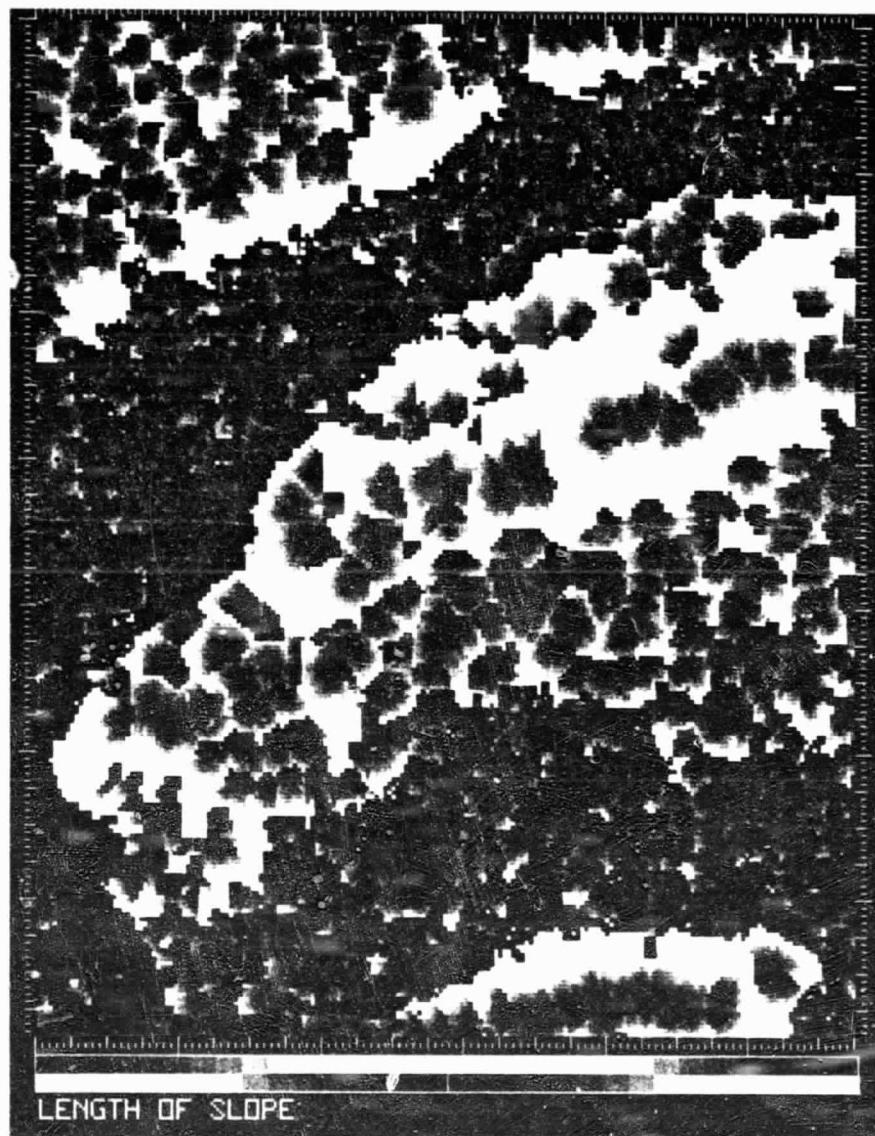


Figure 8 Registered length of slope image generated from DEM. Dark tones represent short slope lengths; light tones represent longer slope lengths.

Slope Gradient

Calculation of slope gradient from the DEM was implemented using a program developed by Professor Jeff Dozier (1979), Department of Geography, University of California, Santa Barbara. Slope angle from a DEM is defined as the gradient of a plane tangent to the centerpoint on a 3 by 3 grid. The four nearest neighbors of each center cell were used to calculate the slope angle of the plane. The output for each cell was generated in radians. The radians were converted to percent slope to coincide with the USLE S coefficient (Figure 9).

For slopes less than nine percent, the relation $(0.43 + 0.30S + .043S^2) / 6.617$, where S equals percent slope, was used to calculate the S factor. For slopes exceeding nine percent the relation $(S/9)^{1.3}$ defined the S coefficient. The S coefficient was calculated in three steps. A binary mask was formed creating two images from the slope image. The first image set slopes greater than nine percent to zero and the first equation was applied. The second image set slopes less than or equal to nine percent to zero and applied the second equation. Finally, the two images were added, generating one slope gradient image with an S value for each cell.

ORIGINAL PAGE IS
OF POOR QUALITY.

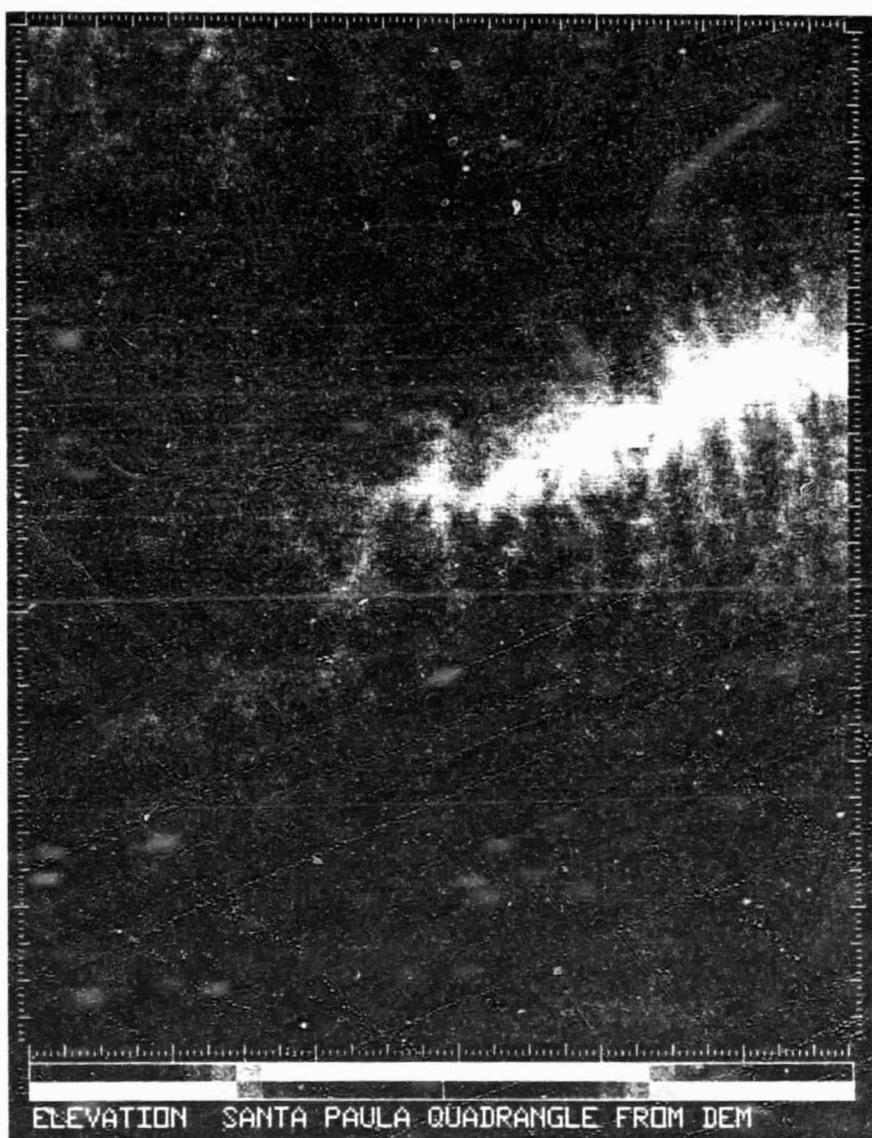


Figure 9 Registered slope gradient image generated from DEM. Dark tones represent lower slope angles; light tones represent steeper slopes.

Crop Management

The Landsat login program VERTSLOG read the raw Landsat tape and transferred the data to disk. VERTSLOG also corrected for skew, aspect ratio, synthetic pixels, mirror scan velocity profile, panorama effect and band to band misregistration. The aspect ratio was set to 1.41, generating 79.9 by 56.7 meter pixels.

The Landsat image was then registered to the Santa Paula seven and one half minute quadrangle. From the Santa Paula and surrounding quadrangles, 1980 color infrared (CIR) imagery at a scale of 1:40,000, and the Landsat sub-scene, nine ground control points were located. The GCP's were mostly highly reflective buildings, or intersections of main highways and railways. The location of the GCP's on the image were tied to their respective location on the quadrangle. A bilinear interpolation was applied using GEOMA to resample the Landsat data to the 60 meter square, 231 line by 192 sample grid. A band 5 display of the registered Landsat sub-scene of the Santa Paula quadrangle is presented in Figure 10.

The geometrically rectified Landsat image was tested against the topographic sheet to determine registration accuracy. Approximately forty randomly selected features on the Landsat image were compared to their locations on the topographic sheet. A majority of the features were in perfect correspondence on the two grids, and none were more than one pixel off.

ORIGINAL PAGE IS
OF POOR QUALITY

51



Figure 10 Registered band 5 Landsat sub-scene of the Santa Paula quadrangle.

The VICAR program USTATS performed an unsupervised clustering algorithm on the June image. USTATS incorporates a modified version of an algorithm suggested by Tyron and Bailey (1972) for clustering large numbers of observations. USTATS initially formed 222 clusters. This was reduced to 188 by USTATS after removing one pixel clusters, and then further reduced to 142 after combining clusters that overlapped by one standard deviation.

The 142 clusters produced by USTATS were reduced to 100 spectral classes using the Numerical Taxonomy System (NTSYS) program package, developed by Rohlf et al. (1974). PDIST, a NTSYS support program, calculated a matrix of standardized Euclidean distances between clusters in the spectral domain. The standardized distance matrix was then input to NTSYS, creating a dendrogram of the clusters. The dendrogram was used to condense the 142 initial clusters to 100 clusters by merging classes or clusters with the classes or clusters to which they were most similar. Some of the small clusters were removed from further processing, while other similar clusters were merged. The VICAR program EDSTATS merged the clusters through a recalculation of new class centroids by pooling old centroids weighted by their number of pixels. New class variances were computed for each channel. The 100 clusters were input to the VICAR program FSTCLSPR, a multispectral classifier using an algorithm which combines parallelepiped and Bayesian maximum likelihood

techniques. All pixels were classified into one of the 100 classes, with 1.8 percent of the data unclassified.

The classes were now labeled. The band 5 Landsat spectral channel was displayed on a Cathode Ray Tube (CRT) utilizing the Grinnel Image Processing System and QDIPS software (Dozier, 1980). The classified image containing 100 classes was overlayed on the band five display. QDIPS facilitated the display of seven colors representing the 100 classes. While the classes were displayed on the CRT, the CIR imagery and the 7.5 minute Santa Paula quadrangle were used to label the statistical clusters to their land use/land cover classes.

After considerable field inspection and air photo analysis, ten land use/land cover classes were derived (Table 1). These classes did not fit a strict Anderson et al. (1976) classification scheme; but instead were designed to distinguish the different effects crop management practices have on erosion potential.

Unfortunately, the classification was not as accurate as anticipated; there was considerable confusion between orchards and natural vegetation. After examination of the 7.5 minute quadrangle and air photos, two conclusions were made:

- (1) No orchard or row crop existed above 800 feet in the Santa Paula quadrangle; and
- (2) Orchards on slopes exceeding 10 degrees were extremely rare.

TABLE 1
LAND USE/LAND COVER CLASSES

Class	Description
Mature Orchard	Citrus or avocado orchards with greater than 50 percent ground cover
Immature Orchard	Citrus or avocado orchards with less than fifty percent ground cover
Row Crop	Any crop grown in rows: including celery, lettuce, beans, etc.
River	Water, sand, rock and riparian vegetation confined to the Santa Clara River channel
Urban	Urban, residential, industrial or transportation land use
Dense Sod	A very dense cover of grass including golf courses and a turf growing outfit
Grass	Mostly Mediterranean annual grasses with very little soft chaparral
Chaparral	Medium to high density soft and hard chaparral
Oak Woodland	Sites with at least a fifty percent oak canopy cover
Barren	Sites mostly devoid of vegetation, not including river areas

Therefore, a binary mask was created using the DEM elevation and slope data identifying all elements greater than 800 feet and all slopes greater than 10 degrees. Classes less than 800 feet in elevation and with a gradient less than 10 degrees retained cluster numbers 1-100. Classes with an elevation exceeding 800 feet and/or with a gradient exceeding 10 degrees were renumbered 101-200. The classified image with 200 classes was now labeled into the ten land use/land cover classes. Class numbers over 100 were generally grass, chaparral, oak woodland or barren. Class numbers under 100 were mature orchard, immature orchard, row crop, dense sod, urban or river classes. The binary mask very accurately separated orchards from natural vegetation.

However, 1.8 percent of the quadrangle remained unclassified. The unclassified regions were generally small isolated clusters of pixels; large groups of unclassified pixels did not exist. SIMPLIFY, a VICAR applications program which removes high frequency components (noise) from a classified or stratified image was applied to the ten class stratified image. The classified pixels were not affected, but the unclassified pixels were assigned to the majority class of the surrounding eight pixels. The classified image is presented in Figure 11.

C coefficients were assigned to each of the land use/land cover classes from the charts and tables provided by the USDA Science and



Figure 11 Registered ten class land use/land cover image generated from Landsat MSS data, and slope and elevation information from DEM.

Education Administration (1978) and supplemental technical notes from the USDA Soil Conservation Davis, California (1977) (Table 2).

TABLE 2
C COEFFICIENTS

Class	C	Description
Mature Orchard	.03	Approximately 60 percent cover
Immature Orchard	.44	Approximately 20 percent cover
Row Crop	.40	Six year row crop sequence with winter cover
River	.00	Not calculated by the SCS
Urban	.00	Not calculated by the SCS
Dense Sod	.01	100 percent cover
Grass	.07	Grass, no appreciable canopy
Chaparral	.01	Brush or bushes, 80 percent cover
Oak Woodland	.01	Trees but no appreciable low brush, 80 percent cover
Barren	.75	No vegetative canopy

Air photo interpretation and field measurements quantified the C coefficients. The derivation of the C coefficient for the orchard classes was based on a combination of air photo analysis and field inspection. The calculations for the row crop and dense sod coefficients were based on the published charts. For the grass and chaparral classes, 100 foot transects for three sites within each class provided an average ground cover. Air photo analysis determined the percent ground cover for the oak woodland, and the barren class C coefficient was obtained from SCS tables. The SCS does not provide guidelines for the calculation of C coefficients for the river and urban classes. Therefore, they were set to zero; predicted soil loss for these sites will be zero tons per acre per year. These classes are not agriculturally important and are located in relatively flat locations.

Conservation Practice

Since conservation measures are not resolvable on Landsat nor available as collateral data, the conservation practice coefficient was not applied in this experiment. The conservation practice factor was set to one, indicating a lack of erosion control techniques. Soil conservation measures are not prevalent in the Santa Paula area. However, locations where erosion reducing techniques are applied will have lower erosion rates than those predicted from this model.

Soil Loss Tolerance

Each soil phase and series has been assigned a soil loss tolerance by the Soil Conservation Service. The soil loss tolerance image was generated in the same manner as the soil erodability; the only difference arose in the final labelling of the soil loss tolerance values. The tolerances ranged from one to five tons per acre per year. Again, the sandy alluvial plain, landslides, igneous rock land, terrace escarpment, sedimentary rock land, fill land, pits and dumps, riverwash and gullied land classes were not assigned soil loss tolerances by the SCS. For the purposes of this study they were assigned tolerance values of zero and applied to 8.8 percent of the soils. Figure 12 displays soil loss tolerances for the Santa Paula quadrangle.

Predicted Soil Loss

The five data planes representing the R,K,L,S and C coefficients were multiplied together based on the Universal Soil Loss equation to derive the A coefficient. The VICAR program F2 can only process a maximum of two input images; therefore, the generation of a predicted soil loss image was a four-step process. First, the $R * K$ images were multiplied. In subsequent steps the $(R * K) * L$, the $(R * K * L) * S$, and finally the $(R * K * L * S) * C$ images were multiplied. A problem with the four step approach is that



Figure 12 Registered soil loss tolerance image generated from digitized Soil Conservation Service soil maps. The lightest tone represents a soil loss tolerance of 1 ton/acre/year; the darkest tone represents a soil loss tolerance of 5 tons/acre/year.

truncation occurs at each stage. This problem is discussed more fully in Chapter Six. A prediction of soil loss in tons per acre per year was calculated for each 60 meter cell in the 231 line by 192 sample grid for the Santa Paula quadrangle (Figure 13).

The final step was the subtraction of the predicted soil loss image from the soil loss tolerance image. The resulting film writer image displays locations where predicted soil erosion exceeds soil loss tolerances (Figure 14).

ORIGINAL PAGE IS
OF POOR QUALITY

62



Figure 13 Registered predicted soil loss based on GIS inputs to the Universal Soil Loss Equation. Lightest tones depict predicted soil loss less than 5 tons/acre/year. Darkest tones represent predicted soil losses exceeding 30 tons/acre/year.

ORIGINAL PAGE IS
OF POOR QUALITY

63



Figure 14 Registered differenced image of predicted soil loss and soil loss tolerance images. Light tones represent locations where the predicted soil loss is less than the soil loss tolerance. Dark tones represent sites where the predicted soil loss exceeds the soil loss tolerance.

CHAPTER 6

ACCURACY AND ANALYSIS

Land Use/Land Cover Classification

A stratified random sample was used to determine the accuracy of the land use/land cover classification. From the VICAR applications program SAMPLE, 25 random samples from each of the ten classes were obtained for accuracy assessment. Line and sample coordinates for each point were listed by the program, 25 for each class. Coordinates of the stratified random sample were transferred to the topographic sheet to aid in the location of the test sites. At this time, SAMPLE did not function properly; the coordinates for some test sites were incorrectly depicted by the SAMPLE routine. These samples were rejected. This was a random process and did not bias the results of the accuracy assessment.

There was easy access to the agricultural portions of the quadrangle in the Santa Clara River Valley and the Las Posas Valley. Many roads traversed the fields and orchards in the lowlands. Here, the method for calculating the accuracy of the classified image was field inspection. It was a relatively simple task to drive near the sample sites and record the land cover.

In the naturally vegetated regions on South Mountain and the foothills north of the Santa Clara River Valley, another method for determining the accuracy of the sample points was necessary. There were fewer roads in this rugged terrain, and some of the ranchers did not permit access to their property. Therefore, air photo interpretation provided the accuracy assessment for the naturally vegetated regions. Using 1:32,000 scale CIR imagery from February 1979, discrimination between the four land cover classes was relatively easy. These classes were grass, chaparral, oak woodland and barren. The large scale of the U-2 imagery, and the response of the vegetation on the infrared film, allowed an accurate assessment of land cover for the naturally vegetated regions.

From the stratified random sample, 84.3 percent of the pixels were correctly classified (Table 3). By weighting each class by its representative area, 87.2 percent of the quadrangle was correctly classified. These are the highest classification accuracies obtained for the Oxnard Plain region. Stow et al. (1980), with only seven classes of interest, achieved 70 percent accuracy in the neighboring Oxnard and Camarillo quadrangles. There was considerable confusion between orchards and chaparral without the elevation and slope information obtained from the DEM. Orchards and chaparral both have large percentages of bare soil interspersed with the vegetation, and Landsat spectral data alone does not differentiate the two classes adequately.

TABLE 3

CLASSIFICATION ACCURACY SANTA PAULA QUADRANGLE

LANDSAT CLASSIFICATION

	mature orchard	im-mature orchard	row crop	river	urban	dense sod	grass	chaparral	oak woodland	barren
CLAS S	22 / 88%	1		1	1					
mature orchard										
immature orchard		21 / 84%			1	1				
row crop		3	13 / 81.3%							
river				18 / 94.7%	1					
urban	1	1	2	4	16 / 66.6%					
dense sod			2			11 / 84.6%				
grass							22 / 91.7%	2		
chaparral	1						2	22 / 91.7%		
oak woodland	1							3	20 / 83.3%	
barren	1	2						1		18 / 81.3%

Pixels Correctly Classified 84.3%

Quadrangle Correctly Classified 87.2%

The accuracies of the land use/land cover classification are relatively consistent for all classes except the urban class. The low accuracy here, 66.6 percent, is due to considerable confusion with the river class. The topographic properties are similar, and the spectral similarities between roads, buildings, and river gravel make discrimination difficult. However, for the study of agricultural erosion, river and urban categories are not classes of interest, and were not included in the final soil loss calculation.

Row crops were on the lower end of classification accuracy, 81.3 percent. There was some confusion between row crops and immature orchards due to the large amount of bare ground found on some row crops at different times of the year, and the large percent of bare soil on the immature orchards. The effects of the confusion are lessened somewhat by the fact that the C coefficients for row crop and immature orchard are .44 and .40, respectively. This similarity reduces the impact of misclassification error on the final soil loss prediction.

Most of the remaining confusion between classes also occurred in classes with relatively similar C coefficients. This is logical, as the C factor is largely based on vegetative cover protecting the soil. Stephens and Cihlar (1981) followed this line of reasoning by correlating the infrared/red ratio with the C coefficient in a study described in Chapter Three. There was some confusion between barren and immature orchards. Both of these classes have

high C coefficients, .74 and .40 respectively. There are very few instances of classes with widely disparate C coefficients displaying confusion. These considerations enhance the overall acceptability of the accuracy value of 87.2 percent.

Collateral Data

Accuracy assessments for the collateral datasets were obtained in the following manner. From the original stratified random sample, ten random samples from each class were used, excluding the river and urban classes. These two classes were eliminated from further investigation. A subsample of 80 sites remained, ten sites for each of the eight classes of interest. The 80 values from the collateral images were tested against the corresponding values for the sources of those images. For example, 80 rainfall values on the rainfall image were tested against the corresponding 80 values on the rainfall map. A problem is that the rainfall, soil erodability and soil loss tolerance are discrete datasets. The slope and length of slope images are continuous datasets. Two different approaches for calculating accuracies were necessary. For the discrete datasets, a standard accuracy percentage scheme was employed. For the continuous datasets, a Pearson product moment correlation coefficient was calculated.

The eighty sites on the rainfall image were tested against the isopluvial polygons drawn on the Santa Paula quadrangle. As

there were only four isopluvial polygons, the accuracy of this dataset was 100 percent. This does not infer that geobased USLE models depiction of the rainfall is 100 percent accurate. It does mean that the geobased model describes the rainfall factor from the USLE completely. Obviously, orographic uplift and subsequent aridity on the lee side of a mountain are not adequately accounted for on the 1:2,000,000 scale NOAA isopluvial maps. However, modeling this effect is beyond the scope of this research.

The soil erodability and soil loss tolerance accuracies were tested in the same manner. The eighty site sample on the soil erodability and soil loss tolerance images were tested against the Soil Conservation Service soil maps. The same 231 by 192 60 meter square grid system was outlined on the SCS soil maps. Line and sample coordinates on the soil property images corresponded directly to the line and sample coordinates on the soil maps. The accuracy for both the soil loss tolerance and soil erodability images tested against the soil maps was 96.25 percent. Each of the soil property datasets had three observations in error. Nearly 1500 soil polygons were manually edited to the appropriate soil loss tolerance and soil erodability values based on the SCS designations. Very few of the polygons were mislabeled in the labelling procedure; the accuracy of the soil erodability and soil loss tolerance images is quite good.

Originally, it was the goal of this study to measure the topographic variables, slope and length of slope, in the field. However, it soon became apparent that this was not possible due to limited access to the naturally vegetated regions. Therefore, the topographic test data was generated from the topographic sheet. The same random sample was applied to test the L and S coefficients for accuracy. The slope was measured from the topographic sheet in the following manner. A distance of 200 feet was measured along the steepest direction through the point in question. This point became the midpoint of the slope. The elevation change was noted by summing the contour lines (20 foot contour interval) for the 200 foot distance, corresponding approximately to the 60 meter DEM cell. The resulting rise over run was the slope for the site.

The slope from the topographic sheet was correlated to the slope generated from Professor Dozier's program. The Pearson product moment Correlation coefficient, R, was .93, significant to the .0001 level. This figure displays very good correlation between the two values. The topographic sheet and the DEM provide independent data sources for the correlation calculation. It is not valid to consider the slope information from the topographic sheet ground truth. The calculation of slope from twenty foot contours on a 1:24,000 scale topographic map is possibly less accurate than slope generated from the DEM. However, the topographic sheet does provide

another source of slope information, and thus can provide a basis for comparison of topographic factors.

For the digital slope length method, the spatial resolution of the DEM is an important consideration in the calculation of slope length. The resampled 60 meter square grid on the DEM does not adequately account for microrelief. Small rises are not discerned on the digital topographic data. Very likely, the DEM exaggerates slope length, missing some of the microrelief which would terminate a slope length. The DEM is sensitive to larger scale topographic factors. These are considerations that must be evaluated in terrain analysis implementing DEM topographic data. The DEM is, however, the highest resolution digital topographic data available for general use by the public.

To test the length of slope, a line was drawn along the steepest slope through each test site. This line followed the steepest slope until the slope became negative. At this point the source of runoff for the initial point was located. On the steeper portions of the quadrangle the slope length ended at a ridge at a peak. On the floodplain, with its more subdued topography, it was more difficult to determine the end of the slope; here, slight variations in elevation concluded the slope length. The Pythagorean distance between the initial point and the end of the slope was calculated as the slope length.

Length of the slope derived from the DEM was correlated to the manually derived length of slope. The Pearson product moment correlation coefficient, R, was .81, significant to the .0001 level. Again, this is simply the correlation of the DEM length of slope and the map-derived length of slope. The map-derived length is certainly not the best method to calculate slope length because the scale on the topographic sheet may be too small for an accurate representation. However, from the relatively high correlation coefficient, it is apparent that the two sets of measurements are reasonably coincident.

Finally, the predicted soil loss derived from the geobased soil loss model was tested against the manually derived coefficients of the USLE. The same set of random samples was applied. The coefficients for the manual USLE were obtained from the rainfall map, the SCS soil maps, slope and length of slope from the topographic sheet, and the groundtruth and air photo analysis of the study area. These coefficients were multiplied together based on the USLE, yielding an A value. The geobased coefficients were also multiplied together using the VICAR program F2, yielding an A value. Because the USLE is a multiplicative function, it is appropriate to transform the geobased and manually based value to their natural logarithms before calculating a correlation coefficient (Li, 1964). The Pearson product moment correlation coefficient, R, of the log transformed data was .91, significant to the .0001 level. This

figure displays very good correlation between the geobased and manually derived USLE soil loss predictions. Classification accuracy from the collateral datasets and predicted soil loss are summarized in Table 4.

TABLE 4
COLLATERAL AND PREDICTED SOIL LOSS DATASET ACCURACIES

Dataset	Accuracy
Rainfall	100.00 Percent
Soil Erodability	96.25 Percent
Length of Slope	.81 Correlation
Slope Gradient	.93 Correlation
Soil Loss Tolerance	96.25 Percent
Predicted Soil Loss	.91 Correlation

Sensitivity Analysis

An informal sensitivity analysis performed on the effects of the five coefficients on the final predicted soil loss is presented in Table 5.

TABLE 5

SENSITIVITY ANALYSIS

Coefficient	Range	Percent Change
R	.50 - .70	140
K	.15 - .43	286
L	1.35 - 2.92	221
S	1.00 - 34.10	3412
C	.01 - .75	7500

The greatest variability was obtained from the slope and crop management coefficients. The misclassification of chaparral ($C = .01$) as barren ($C = .75$) created an error in the A factor of 7500 percent. However, this error occurred only once in the sample analyzed. Slope is a very sensitive coefficient of the USLE. The slope factor depicted by the DEM is fairly accurate, and error within this calculation is unlikely to extend the full range of the S value. Length of slope and the soil erodability factors are moderately sensitive to variation within their respective coefficients. The soil image was quite accurate and does not account for much error in the A calculation. The length of slope is the least accurate of the datasets. Fortunately, it has the second lowest sensitivity of the five coefficients, reducing the impact of miscalculation of the L coefficient on the predicted soil loss. The rainfall factor is quite stable, and probably does not add error to the A calculation.

One final test was undertaken as a check on the VICAR processing sequence. The F2 program, which performs arithmetic functions on images, can process a maximum of two images at a time. Multiplication of the five datasets of the USLE was accomplished in four steps as described in the previous chapter. Unfortunately, only integer values are allowed as output from F2. From the series of four multiplications, some error may have been introduced by truncation. A test was conducted to determine if this was the case.

The five coefficients of the geobased model for the same test sample (80 values) were manually multiplied based on the USLE. The results were tested against the geobased automated A value. The Pearson product moment correlation Coefficient R value was .98, significant to the .0001 level. Obviously, little error was introduced by the truncation of values during the F2 processing sequence.

Analysis of Predicted Soil Loss

The images of predicted soil loss and soil loss versus soil loss tolerance show that the orchards and row crops along the floodplain do not display soil loss problems. Soil losses here are less than five tons per acre per year and, for the most part, are less than the soil loss tolerances. There are several locations in the flood plain where predicted soil losses exceed soil loss tolerances. In the northeastern portion of the Las Posas Valley there is a region of steep slopes planted in row crops. These sites display an increased predicted soil loss. Along the western portion of the Santa Clara River, row crops planted on soils with soil loss tolerances of one ton per acre per year exceed their soil loss tolerance. Immature orchards, mostly avocado, extending into the canyons of the mountainous areas display high predicted soil loss rates. The problem is not as widespread on the Santa Paula quadrangle as was previously theorized.

Slope and crop management appear to be the strongest indicators of soil loss problems. On the fairly level sites soil loss is relatively minor; on the steeper and longer slope sites soil loss is considerable. The grass sites often have minimal vegetation cover, and soil loss is potentially high. Predictions consistently exceed soil loss tolerances in the grass areas on South Mountain and in the foothills north of the Santa Clara River Valley. Chaparral ($C = .01$) is superior to grass ($C = .07$) in protecting the ground from rainfall erosion losses. The prevalent fires on South Mountain remove the chaparral cover; a grass canopy returns. This succession causes an increase in predicted erosion for the mountainous areas.

Several trends are apparent from analysis of the predicted soil loss image:

- (1) Soil loss is not a problem on the floodplains and valleys, except with row crops on soils with low soil loss tolerances or on the steeper slopes;
- (2) Immature avocado orchards display high predicted soil losses in the canyon sites, however, the problem is not extensive; and,
- (3) The steeper mountainous locations are serious soil loss areas, especially those areas exhibiting a grass canopy.

Thus, the geobased soil loss model accurately represents the manual Universal Soil Loss Equation. Although imperfect, this

soil loss information system has potential to inventory large areas for predicted soil loss with a savings in both time and money over conventional ground sampling (Hanuschak et al. 1979). This model is probably best applied as a prescreening mechanism to identify major areas of soil loss, in which a user is more concerned with the relative amounts and the spatial extent of the predicted soil losses. The geobased model certainly provides this information.

CHAPTER 7

CONCLUSION

This thesis research has demonstrated the capability of a geographic information system including Landsat and collateral data to map potential soil loss using the Universal Soil Loss Equation. This geobased soil loss information system accurately depicted the Universal Soil Loss Equation for a 7.5 minute quadrangle, an agricultural region exceeding 100 square miles. The prediction of soil erosion for a small site is useful, but the ability to project this prediction to a larger area provides a much greater perspective of soil loss problems to resource managers.

As an example, Mugu lagoon, at the mouth of Revlon Slough, is experiencing serious sedimentation problems. Sediment is transported from the upper watershed; much of it is from the Santa Paula quadrangle. The predicted soil loss image clearly shows that locations of steeper slopes under grass canopy are the major contributing sources for soil lost from this area. Resource managers can implement this information by introducing a more protective plant cover in this area, such as chaparral or sod-forming grasses. The introduction of immature orchards in canyons extending into the mountainous areas is increasing soil loss problems. Conservation practice techniques such as contour terracing can reduce the

erosion hazard. Another alternative is a zoning ordinance prohibiting or placing restrictions on the planting of orchards in critical areas identified by the geobased soil loss model. The decision is the resource manager's; the data are provided by the VICAR/IBIS soil loss model.

The intermediate images developed and employed in the research display ancillary information in an effective manner. The soil erodability and soil loss tolerance images reduce the soil maps from approximately 1500 soil mapping units to more comprehensive five- to eight-class images. These simplifications are helpful because the distribution of the soil erodability and soil loss tolerances for soil classes is not obvious on complex soil maps. The length of slope image graphically depicts the distance from each pixel to the source of runoff for that point; topographic sheets do not provide as clear a representation as the image. The land use/land cover classification image displays information in a form more interpretable than Ventura County land use maps. And finally, the VICAR/IBIS data base is a set of overlain digital rasters, allowing versatility in processing and display.

Additional information can be generated from the intermediate data sets. Since the borders of the soil polygons exists as an IBIS file, soil properties, such as permeability and PH, which have been calculated by the SCS for the soil mapping units, can easily be displayed as images. Permeability influences infiltration

rates and the magnitude of ground water recharge; soil PH can provide an assessment of crop suitability for a site. From the slope gradient image, discrete ranges in slope can be generated (i.e., five to ten, ten to fifteen, fifteen to twenty percent slope, etc.), and the location and the percentage of each slope increment can be determined. Thus, there is considerable potential for an automated approach to data handling and display; the limitation is the imagination and the creativity of the user of the geographic information system.

Statistics from the intermediate data sets are useful as well. Histograms for the soil erodability and soil loss tolerance images have been generated from this work. This information can help the Soil Conservation Service as well as county planners, as such data are currently unavailable from other sources. A histogram of the land use/land cover classification was also produced. Again, percentages of land use/land cover classes for this quadrangle are unavailable from the Ventura County Planning office.

The mapping and display of watershed characteristics from DEM also shows great promise. Typical parameters derived from digital terrain data include slope gradient and aspect. In this work we have developed length of slope, which is used in the USLE, but there are other applications for this variable as well. An indication of the amount of moisture available for vegetation at a specific site can be obtained from the length of slope. Shorter

slope lengths have less runoff and they are, therefore, less likely to become completely saturated. Sites with longer slope lengths have a greater source for overland flow providing an additional moisture availability. For runoff routing schemes, the prediction of volume of runoff through a point can be facilitated by length of slope information. Other topographic information, such as the delineation of watershed divides, can also be obtained from the DEM. Thus, a package of watershed variables derived from the DEM has considerable potential for watershed analysis.

Extension of this research might involve a prediction of soil erosion rates for an entire watershed. From this predicted soil loss, a sediment routing scheme similar to Kling's (1973) could be developed. A predicted soil loss map could be generated, as well as a map depicting potential sediment accumulations for the watershed. Designed within the context of a GIS, these simulations would enhance a resource manager's ability to assess erosion problems for an agricultural region.

For this study, the intermediate datasets were accurate and flexible in their representation of environmental information. The predicted soil loss and the soil loss versus soil loss tolerance images correctly identified predicted soil loss problem areas for the Santa Paula 7.5 minute quadrangle. The strength of a geographic information system lies in the accuracy of the input data, and the ability of the system to process and display the data in a suitable

fashion for resource analysis. In this, it is the authors' opinion that the VICAR/IBIS soil loss information system succeeded on all counts.

The following is a summary of conclusions drawn from this study. Listed below are conclusions from the accuracy tests.

- (1) The correlation, R, of the predictions of the VICAR/IBIS soil loss model and the manual USLE values was .91 (log transform), thus explaining most of the variance of the relationship. The system was highly effective in identifying areas of critical soil loss with this degree of error.
- (2) The Landsat classification accuracy using a collateral binary mask derived from slope and elevation information from the DEM was 84.3 percent, with an estimated 87.2 percent of the quadrangle correctly classified. This accuracy is very good; most of the misclassification was from classes with similar C coefficients. Accuracies were considerably lower without topographic information. The registration of the Landsat to the Santa Paula quadrangle with nine GCP's was excellent, employing a bilinear interpolation resampling approach.
- (3) The DEM is an accurate high resolution data source. Precise registration of the DEM to the Landsat from the four corners of the DEM was achieved.

- (4) Slope gradient from the DEM as tested against the topographic map displayed a correlation coefficient, R, of .93, a very good correlation.
- (5) Length of slope was the least accurate dataset. Correlation to the test data provided an R coefficient of .81. More work is necessary on the length of slope routine. It is unclear if the scale of the DEM permits a sufficiently precise depiction of the microrelief necessary for a length of slope calculation.
- (6) Both the soil and rainfall images were input to the system with little error.
- (7) Processing using the VICAR F2 sequence was acceptably accurate in spite of possible roundoff errors. The five-variable multiplication as tested against a manual multiplication correlated with a coefficient, R, of .98.
- (8) Sensitivity analysis showed that slope and crop cover were potentially the most sensitive factors of the five USLE coefficients in predicting soil loss.
- (9) The most serious soil loss problems are in areas with steeper slopes under grass cover. Immature orchards in the foothills of the mountains display high predicted soil losses; however, these sites are not extensive.

General conclusions from this study are as follows.

- (1) Incorporating topographic data in Landsat classification is an extremely efficient method to improve classification accuracies. However, a thorough knowledge of the study area and of environmental processes affecting land use/land cover are prerequisites for success in this endeavor.
- (2) Watershed characteristics can be effectively derived from DEM data. This study implemented slope and length of slope information, as well as elevation data. Length of slope information can be used in applications other than as the USLE variable, and watershed divide information can also be obtained from DEM topographic data.
- (3) Digitized soil maps are a powerful information source. Once the borders are coded, attributes other than erodability and soil loss tolerance, such as permeability and PH, can be incorporated into a GIS.
- (4) Useful statistics from soil maps, topographic data and land cover data were obtained. The percentage of slopes within a certain range, the proportion of specific soil erodability and soil loss tolerances for soils, and the percentage of land covered in certain land use/land cover categories can be easily calculated.

- (5) The 60 meter square 231 line by 192 sample grid for all of the datasets, as well as the topographic sheet and soil maps, was an effective coordinate system. Ground survey and accuracy assessments were greatly facilitated by the use of this grid. The grid also aided the registration of the Landsat and the DEM to the Santa Paula quadrangle.
- (6) The use of topographic maps for accuracy verification in this study is questionable. The scale of the topographic sheets may not adequately depict slope or length of slope. However, field data for these factors were not obtainable.
- (7) All of the data used in this study are available in the United States, except the DEM. Defense Mapping Agency topographic data may be substituted for the DEM, but the DMA data are not as accurate. In time, the DEM will be more widely available.
- (8) VICAR/IBIS proved to be a powerful, flexible, and well integrated image processing and geographic information system. Problems with VICAR/IBIS were with the two-image F2 capability, and the POLYSCRB routine, which randomly assigns border pixels. However, VICAR/IBIS effectively processed all of the datasets according to user needs.
- (9) A geographic information system based Universal Soil Equation can be applied to a wide range of agricultural regions, similar in climate to the continental United States. Application to forested watersheds is not recommended because the

USLE was developed to predict erosion for agricultural land. Application of this model to other climatic zones, such as tropical environments, is subject to further research on soil erosion processes for these specific areas.

REFERENCES AND SELECTED BIBLIOGRAPHY

- Anderson, J.R., E.E. Hardy, J.T. Roach and R.E. Witmer, 1976, "A Land Use and Land Cover Classification for Use With Remote Sensing Data," U.S.G.S. Journal of Research, Vol. 5, No. 2, pp. 143-153.
- Andre, J.E. and H.W. Anderson, 1976, "Variation of Soil Erodability with Geology, Geographic Zone, Elevation, and Vegetation Type in Northern California Wildlands," Journal of Geophysical Research, Vol. 66, No. 10, pp. 3351-3356.
- Berger, Z. and J.R. Jensen, 1980, "Modelling Soil Loss and Flood Potential Due to Urbanization in Humid Subtropical South-eastern Environments," Proceedings of the Fourteenth International Symposium on Remote Sensing of Environment, pp. 1057-1068.
- Brady, N.C., 1974, The Nature and Property of Soils, The MacMillan Publishing Company, 639 p.
- Bryant, N.A. and A.A. Zobrist, 1976, "IBIS: A Geographic Information System Based on Digital Image Processing and Image Raster Datatype," Second Annual Symposium on Machine Processing of Remotely Sensed Data: Laboratory for Applications of Remote Sensing, Purdue University, pp. 1A1-1A7.
- Carson, M.A. and M.J. Kirkby, 1972, Hillslope Form and Process, Cambridge University Press, 475 p.
- Castleman, K.R., 1979, Digital Image Processing, Prentice-Hall Inc., 429 p.
- Collins, S.H., 1975, "Terrain Parameters Directly From a Digital Terrain Model," Canadian Surveyor, Vol. 29, No. 5, pp. 507-518.
- Degani, A., L. Lewis and B. Downing, 1979, "Interactive Computer Simulation of the Spatial Process of Soil Erosion," Professional Geographer, Vol. 31, No. 2, pp. 184-190.
- Dozier, J., 1980, Use of Environmental Satellite Data for Input to Energy Balance Snowmelt Models, Final Report NOAA Grant 04-8-M0, 82 p.

Dozier, J. and S.I Outcalt, 1979, "An Approach Toward Energy Balance Simulation Over Rugged Terrain," Geographical Analysis, Vol. 2, No. 1, pp. 65-85.

Estes, J.E., L.R. Tinney and D.A. Stow, 1979, Conceptual Design and Preliminary Demonstration of a Prime Agriculture Land Component of an Integrated Remote Sensing System, California Integrated Remote Sensing Systems, 30 p.

Evans, W. and G. Kalkanis, 1976, "Use of the Universal Soil Loss Equation in California," Soil Erosion Prediction and Control: Proceedings of the Soil Conservation Society of America, pp. 31-40.

Foster, G.R. and L.D. Meyer, 1972, "Transport of Soil Particles by Shallow Flow," Transactions American Society of Agricultural Engineers, Vol. 15, No. 1, pp. 99-102.

Hallada, W.A., F.C. Mertz, L.R. Tinney, M.J. Cosentino and J.E. Estes, 1981, "Flexible Processing of Remote Sensing Data Through Integration of Image Processing and Geobased Information Systems," Proceedings of the Fifteenth International Symposium on Remote Sensing of Environment, (in press).

Hanuschak, G., R. Sigman, M. Craig, M. Ozga, R. Luebbe, P. Cook, D. Kleweno and C. Miller, 1979, "Crop-Area Estimates From Landsat: Transition From Research and Development to Timely Results," Fifth Annual Symposium on Machine Processing of Remotely Sensed Data: Laboratory for Applications of Remote Sensing, Purdue University, pp. 86-94.

Hutchinson, C.F., 1982, "Techniques for Combining Landsat and Ancillary Data for Digital Classification Improvement," Photogrammetric Engineering and Remote Sensing, Vol. 48, No. 1, pp. 123-130.

Joyce, A.T., 1979. Final Report of the Natural Resources Inventory System ASVT Project, NASA Technical Memorandum, NASA TM 58211, 129 p.

Junkin, Bobby G., 1979, A Method for the Processing and Analysis of Digital Terrain Elevation Data, NASA National Space Technology Laboratories, Report #177, 26 p.

- Keller, E.A., D.L. Johnson, T.K. Rockwell, M.N. Clark and G.R. Dembroff, 1981, Quaternary Stratigraphy, Soil Geomorphology, Chronology and Tectonics of the Ventura, Ojai and Santa Paula Areas, Western Transverse Range, California, Friends of the Pleistocene Field Guide, 159 p.
- Kling, G.F. and G.W. Olson, 1974, "The Sediment Transport Computer Model," Cornell Agronomy Memo.
- Li, C.C., 1964, Introduction to Experimental Statistics, McGraw-Hill Book Company, 460 p.
- Logan, T.L., 1978, Western Forest Species Classification from Multidate Landsat and Digital Terrain Data, Master's Thesis, Department of Geography, University of California, Santa Barbara, 82 p.
- Logan, T.L., 1981, "A Data Base Approach for Prediction of Deforestation-Induced Mass Wasting Events," Proceedings American Society of Photogrammetry, pp. 197-211.
- Meeuwig, R.O., 1971, "Soil Stability on High Elevation Rangeland in the Intermountain Area," U.S. Department of Agriculture, Forest Service Research Paper, INT-94.
- Mertz, F.C., M.J. MacLennan, T.A. Streich, L.R. Tinney, 1982, JSC VICAR/IBIS Primer, Geography Remote Sensing Unit, University of California, Santa Barbara, (in press).
- Miller, J.F., R.H. Frederick and R.J. Tracey, 1973, Precipitation Frequency Atlas of the Western United States, National Oceanic and Atmospheric Administration, 71 p.
- Moik, J.G., 1980, Digital Image Processing of Remotely Sensed Images, NASA SP-431, Scientific and Technical Information Branch, 330 p.
- Morgan, K.M., D.R. Morris-Jones, G.B. Lee and R.W. Kiefer, 1979, "Cropping Management Using Color and Color Infrared Aerial Photographs," Photogrammetric Engineering and Remote Sensing Vol. 45, No. 6, pp. 769-774.
- Morgan, K.M., D.R. Morris-Jones, G.B. Gerhard and R.W. Kiefer, 1980, "Airphoto Analysis of Erosion Control Practices," Photogrammetric Engineering and Remote Sensing, Vol. 46, No. 5, pp. 637-642.
- Musgrave, G.W., 1947, "The Quantitative Evaluation of Factors in Water Erosion, A First Approximation," Journal of Soil and Water Conservation, Vol. 2, No. 3, pp. 133-138.

NOAA Environmental Data and Information Service, 1978, Climatological Data: California, National Oceanic and Atmospheric Administration, 28 p.

Norris, R.M. and R.W. Webb, 1978, Geology of California, John Wiley and Sons, Inc., 365 p.

Olson, T.C. and W.H. Wischmeier, 1963, "Soil Erodability Evaluations for Soils on the Runoff and Erosion Stations," Soil Science Society of America Proceedings, Vol. 27, pp. 590-592.

Pacific Southwest Inter-Agency Committee, 1968, Factors Affecting Sediment Yield and Measures for the Reduction of Erosion and Sediment Yield, Water Resources Council, 10 p.

Pacific Southwest Inter-Agency Committee, 1971, Comprehensive Framework Study, California Region Watershed Management, Water Resources Council, Appendix VIII, 156 p.

Patterson, F.A. and P.M. McAdams, 1980, The Use of Landsat MSS Data to Produce Erosion Hazard Potential Maps, NASA National Space Technology Laboratories, 18 p.

Rodriguez, B., 1978, "Application of Landsat and Skylab Imagery in Mexico: Detection of Erosion and Forest Damage," Proceedings of the Twelfth International Symposium on Remote Sensing of Environment, pp. 1609-1616.

Rohlf, F.J., J. Kishpaugh and D. Kirk, 1974, NT-SYS Numerical Taxonomy System of Multivariate Statistical Programs, New York State University, Stony Brook and Quantra Corporation.

Sabins, F.F., 1978, Remote Sensing Principles and Interpretation, W.H. Freeman and Company, 426 p.

Scott, K.M. and P. Williams, 1979, Erosion and Sediment Yields in the Transverse Ranges, United States Geological Survey Professional Paper 1030, 38 p.

Short, N., P. Lowman, S. Frieden and W. Finch, 1976, Mission to Earth: Landsat Views the World, NASA SP 360.

Simonett, D.S., T.R. Smith, W. Tobler, D.G. Marks, J.E. Frew and J.C. Dozier, 1978, Geobase Information System Impacts on Space Image Formats, Santa Barbara Remote Sensing Unit Technical Report #3, 122 p.

- Singer, M.J., G.L. Huntington and H.R. Sketchley, 1976, "Erosion Prediction on California Rangeland: Research Developments and Needs," Soil Erosion Prediction and Control, Proceedings Soil Conservation Society of America, pp. 143-151.
- Stevens, P.R. and J. Cihlar, 1981, "The Potential of Remote Sensing for Monitoring Soil Erosion on Cropland," Proceedings of the Fifteenth International Symposium on Remote Sensing of Environment, (in press).
- Stow, D.A., 1978, Analysis of Landsat/MSS and Digital Terrain Tape Data in a Geobase Information Systems Context: Ventura County Study Area, Masters Thesis, Department of Geography, University of California, Santa Barbara, 167 p.
- Stow, D.A., 1980, "Deriving Land Use/Land Cover Statistics from Landsat: A Study of Prime Agricultural Land," Proceedings of the Fourteenth International Symposium on Remote Sensing of Environment, pp. 1227-1238.
- Strahler, A.H., J.E. Estes, P.F. Maynard, F.C. Mertz and D.S. Stow, 1980, "Incorporating Collateral Data in Landsat Classification and Modelling Procedures," Proceedings of the Fourteenth International Symposium on Remote Sensing of Environment, pp. 1009-1026.
- Strahler, A.H., 1981, "Stratification of Natural Vegetation for Forest and Rangeland Inventory Using Landsat Digital Imagery and Collateral Data," International Journal of Remote Sensing, Vol. 2, No. 1, pp. 15-41.
- Strahler, A.N., 1957, "Quantitative Analysis of Watershed Geomorphology," Transactions American Geophysical Union, Vol. 38, No. 6, pp. 913-920.
- Swain, P.H. and S.M. Davis, 1978, Remote Sensing: The Quantitative Approach, McGraw-Hill Inc., 396 p.
- Swanson, D.N., F.J. Swanson and D. Rosgen, 1980, "Soil Mass Movement," An Approach to Water Resources Evaluation Non-Point Sources: Silviculture, USDA/EPA Environmental Research Laboratories, Athens, Georgia.

Tinney, L.R. and J.E. Estes, 1981, "Agricultural Land Use Change Detection," Proceedings of the Seventh W.T. Pecora Symposium on Remote Sensing: An Input to Geographic Information Systems in the 80's, (in press).

Tryon, R. and D. Bailey, 1972, Cluster Analysis, McGraw-Hill, pp. 147-150.

Tucker, C.J. and L.D. Miller, 1977, "Soil Spectra Contribution to Grass Canopy Spectral Reflectance," Photogrammetric Engineering and Remote Sensing, Vol. 42, No. 6, pp. 721-726.

USDA Agricultural Research Service, 1961, "A Universal Equation for Predicting Rainfall-Erosion Losses," ARS 22-66.

USDA Science and Education Administration, 1978, "Predicting Rainfall Erosion Losses: A Guide to Conservation Planning," Agricultural Handbook #537, 58 p.

USDA Soil Conservation Service, Davis, California, 1977, Guides for Erosion and Sediment Control, United States Department of Agriculture, 32 p.

USDA Soil Conservation Service, Ventura County, 1969, Summary of Soil Characteristics and Qualities, Ventura Area, California, United States Department of Agriculture, 104 p.

USGS National Cartographic Information Center, 1980, Users Guide to Digital Elevation Models, United States Geological Survey, (draft).

Varnes, D.J., 1958, "Landslides Types and Processes," Landslides and Engineering Practice: Highway Research Board, Special Report #29, pp. 20-47.

Ventura County Department of Agriculture, 1980, Agricultural Crop Report, 10 p.

Wischmeier, W.H., 1960, "Cropping Management Factor Evaluations for a Universal Soil Loss Equation," Soil Science of America Proceedings, Vol. 23, pp. 322-326.

Wischmeier, W.H., 1962, "Rainfall Erosion Potential," Agricultural Engineering, Vol. 43, No. 4, pp. 212-215.

Wischmeier, W.H., C.B. Johnson and B.V. Cross, 1971, "A Soil Erodability Nomograph for Farmland and Construction Sites," Journal of Soil and Water Conservation, Vol. 26, No. 5, pp. 189-193.

Wischmeier, W.H., D.D. Smith and R.E. Uhland, 1958, "Evaluation of Factors in the Soil Loss Equation," Agricultural Engineering, Vol. 39, No. 8, pp. 458-462.

Wischmeier, W.H. and D.D. Smith, 1965, "Predicting Rainfall Erosion Losses from Cropland East of the Rocky Mountains," U.S.D.A. Agricultural Research Service, 45 p.

Woodcock, C.E., A.H. Strahler and T.L. Logan, 1980, "Stratification of Forest Vegetation for Timber Inventory Using Landsat and Collateral Data," Proceedings of the Fourteenth International Symposium on Remote Sensing of Environment, pp. 1769-1787.

Young, A., 1972, Slopes, Oliver and Boyd Inc., 288 p.

Zingg, A.W., 1940, "Degree and Length of Slope as it Affects Soil Loss in Runoff," Agricultural Engineering, Vol. 21, No. 2, pp. 59-64.